

ALPHA C. CHIANG concluiu seu doutorado na Columbia University em 1954, após terminar seus estudos superiores na St. John's University (Shanghai, China) em 1946 e fazer mestrado na University of Colorado em 1948. Em 1954, juntou-se ao corpo acadêmico da Denison University em Ohio, onde assumiu a chefia do Departamento de Economia em 1961. De 1964 em diante, lecionou na University of Connecticut, onde, após 28 anos, tornou-se Professor Emérito de Economia em 1992. É também Professor Visitante do New Asia College, da Chinese University of Hong Kong, da Cornell University, da Lingnan University em Hong Kong e da Helsinki School of Economics and Business Administration. Entre suas publicações está um outro livro sobre economia matemática: *Elements of Dynamic Optimization*, Waveland Press, Inc., 1992. Entre as honrarias que recebeu figuram prêmios da Ford Foundation e associação à National Science Foundation, eleição para a presidência da Ohio Association of Economists and Political Scientists, 1963–1964, e citação na publicação *Who's Who in Economics: A Biographical Dictionary of Major Economists 1900–1994*, MIT Press.

KEVIN WAINWRIGHT é membro do corpo acadêmico do British Columbia Institute of Technology em Burnaby, B.C., Canadá. Desde 2001, ocupa o posto de presidente da associação de docentes e de chefe do programa de Administração de Empresas. Estudou na Simon Fraser University em Burnaby, B.C., Canadá, e continua a lecionar no Departamento de Economia dessa Universidade. Especializou-se em teoria microeconômica e economia matemática.

Consulte nosso catálogo completo e últimos lançamentos em: www.campus.com.br



Uma empresa Elsevier
www.campus.com.br



Este livro aborda os seguintes tipos principais de análise econômica: estática (análise de equilíbrio), estática comparativa, problemas de otimização, dinâmica e otimização dinâmica. Nessas análises são introduzidos os seguintes métodos matemáticos: álgebra matricial, cálculo diferencial e integral, equações diferenciais, equações de diferença e teoria do controle ótimo. Este livro servirá também como leitura suplementar em cursos de Microeconomia, Macroeconomia e crescimento e desenvolvimento econômico.

A presente edição contém diversas modificações significativas. O material sobre programação matemática agora é apresentado mais cedo, em um novo Capítulo 13, intitulado “Tópicos Adicionais de Otimização”. Esse capítulo tem dois temas principais: otimização com limitações de desigualdade e o teorema do envelope. No primeiro tema, as condições de Kuhn-Tucker são desenvolvidas de um modo muito semelhante ao que foi empregado na edição anterior. Contudo, o tópico foi aprimorado com diversas novas aplicações econômicas, entre elas a determinação de preço de carga de pico e o racionamento de consumo. O segundo tema refere-se ao desenvolvimento do teorema do envelope, à função de valor máximo e à noção de dualidade. Aplicando o teorema do envelope a vários modelos econômicos, obtemos resultados importantes como a identidade de Roy, o lema de Shepard e o lema de Hotelling.

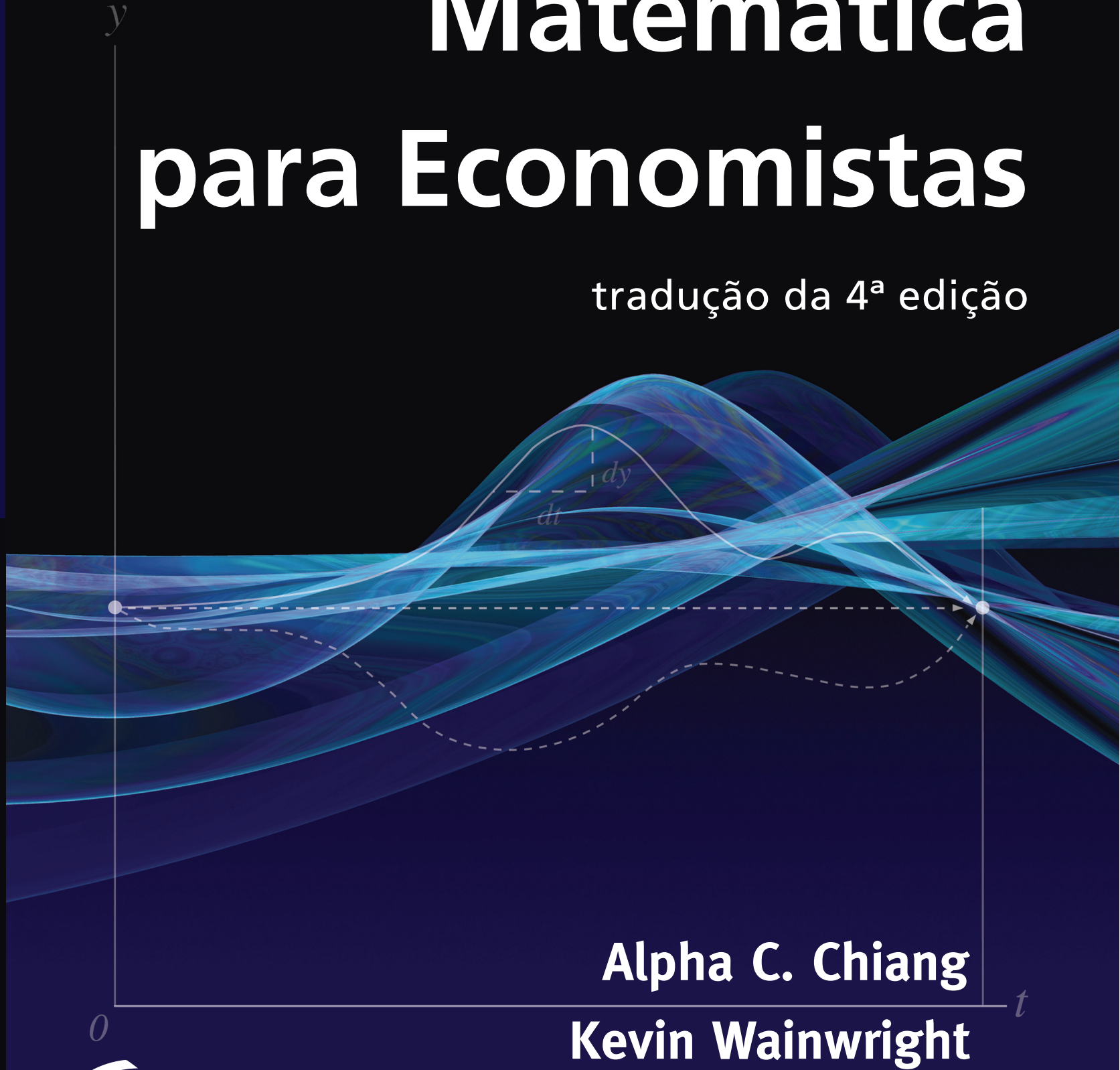
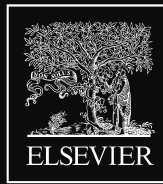
A segunda maior adição a esta edição é um novo Capítulo 20 sobre a teoria do controle ótimo. O propósito deste capítulo é apresentar ao leitor o básico do controle ótimo e demonstrar como ele pode ser aplicado à economia, incluindo exemplos da economia de recursos naturais e da teoria do crescimento ótimo.

Além desses dois novos capítulos, há diversos acréscimos e refinamentos nesta edição. No Capítulo 3, ampliamos a discussão da resolução de equações polinomiais de grau mais alto por fatoramento. No Capítulo 4, foi acrescentada uma nova seção sobre cadeias de Markov. E, no Capítulo 5, introduzimos a verificação da classificação de uma matriz por meio de uma matriz escalonada e a condição de Hawkins-Simon em conexão com o modelo de insumo-produto de Leontief. No que diz respeito a aplicações econômicas, foram acrescentados muitos novos exemplos, e algumas das aplicações existentes foram aprimoradas. Foi incluída uma versão linear do modelo IS-LM, e uma forma mais geral do modelo foi ampliada para abranger tanto uma economia aberta quanto uma fechada, demonstrando, assim, uma aplicação muito mais rica da estática comparativa a modelos de função geral. Entre outros acréscimos estão uma discussão da utilidade esperada e da preferência de riscos, um modelo de maximização de lucros que incorpora a função de produção de Cobb-Douglas e um problema de escolha intertemporal de dois períodos. Por fim, os problemas e exercícios de fixação foram revisados e aumentados para dar aos alunos uma oportunidade maior de aprimorar suas habilidades.



Chiang
Wainwright

Matemática
para Economistas



O livro de Alpha Chiang, desta vez em conjunto com Kevin Wainwright, é uma excelente introdução à Economia Matemática. Os autores apresentam os conceitos básicos da álgebra linear e do cálculo ao longo do livro, em meio a uma grande variedade de aplicações à economia.

Além disso são apresentados muitos tópicos de grande importância para economistas, como o estudo de máximos e mínimos tanto sem condicionamento quanto condicionados. O livro contém também uma introdução às equações diferenciais, à otimização dinâmica e à teoria do controle ótimo.

Nesta edição, os exemplos foram atualizados e foram adicionadas novas aplicações. O presente texto é uma valiosa adição à bibliografia, tão escassa, de Economia Matemática.

Rafael José Iorio Jr.
Pesquisador Titular do Instituto Nacional de Matemática Pura e Aplicada (IMPA)



Matemática para Economistas



Preencha a **ficha de cadastro** no final deste livro
e receba gratuitamente informações
sobre os lançamentos e as promoções da
Editora Campus/Elsevier.

Consulte também nosso catálogo
completo e últimos lançamentos em
www.campus.com.br

Matemática para Economistas

tradução da 4ª edição

Alpha C. Chiang
Kevin Wainwright

Consultoria Editorial
Gerson Lachtermacher
Professor adjunto FCE/UERJ e EBAPE/FGV

Tradução
Arlete Simille Marques

Revisão Técnica
Rafael José Iorio Jr.
*Pesquisador Titular do Instituto Nacional
de Matemática Pura e Aplicada (IMPA)*



2ª Tiragem



Do original:

Fundamental Methods of Mathematical Economics

Tradução autorizada do idioma inglês da edição publicada por MacGraw-Hill/Irwin – McGraw-Hill Companies, Inc.

Copyright © 2005 McGraw-Hill Companies, Inc.

© 2005, Elsevier Editora Ltda.

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998.

Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.

Copidesque:

Cláudia Mello Belhassof

Editoração Eletrônica:

Estúdio Castellani

Revisão Gráfica:

Marília Pinto de Oliveira e Marco Antonio Correa

Capa: Kami Carter

Projeto Gráfico

Elsevier Editora Ltda.

A Qualidade da Informação.

Rua Sete de Setembro, 111 – 16º andar

20050-006 Rio de Janeiro RJ Brasil

Telefone: (21) 3970-9300 FAX: (21) 2507-1991

E-mail: info@elsevier.com.br

Escritório São Paulo:

Rua Quintana, 753/8º andar

04569-011 Brooklin São Paulo SP

Tel.: (11) 5105-8555

ISBN 13: 978-85-352-1769-8

ISBN 10: 85-352-1769-8

Edição original: ISBN 0-07-010910-9

Nota: Muito zelo e técnica foram empregados na edição desta obra. No entanto, podem ocorrer erros de digitação, impressão ou dúvida conceitual.

Em qualquer das hipóteses, solicitamos a comunicação à nossa Central de Atendimento, para que possamos esclarecer ou encaminhar a questão.

Nem a editora nem o autor assumem qualquer responsabilidade por eventuais danos ou perdas a pessoas ou bens, originados do uso desta publicação.

Central de atendimento

Tel.: 0800-265340

Rua Sete de Setembro, 111, 16º andar – Centro – Rio de Janeiro

e-mail: info@elsevier.com.br

site: www.campus.com.br

Sobre a Capa: O gráfico da Figura 20.1 na página 611 ilustra que a menor distância entre dois pontos é uma linha reta. Essa imagem foi escolhida para ilustrar a capa do livro porque uma verdade assim tão simples exige uma das técnicas mais avançadas descritas aqui.

CIP-Brasil. Catalogação-na-fonte.
Sindicato Nacional dos Editores de Livros, RJ

C454m

Chiang, Alpha C., 1927-

Matemática para economistas / Alpha C. Chiang, Kevin Wainwright ;

tradução de Arlete Simille Marques. – Rio de Janeiro : Elsevier, 2006

– 2ª Reimpressão.

il.

Tradução: Fundamental methods of mathematical economics

Anexos

Inclui bibliografia

ISBN 85-352-1769-8

1. Economia matemática. I. Wainwright, Kevin. II. Título.

05-3940.

CDD 330.1543

CDU 330.4

Para Emily, Darryl e Tracey

– *Alpha C. Chiang*

Para Skippy e Myrtle

– *Kevin Wainwright*

Os Autores

Alpha C. Chiang concluiu seu doutorado na Columbia University em 1954, após terminar seu bacharelado na St. John's University (Shanghai, China) em 1946 e fazer mestrado na University of Colorado em 1948. Em 1954, juntou-se ao corpo acadêmico da Denison University em Ohio, onde assumiu a chefia do Departamento de Economia em 1961. De 1964 em diante, lecionou na University of Connecticut, onde, após 28 anos, tornou-se Professor Emérito de Economia em 1992. Foi também Professor Visitante do New Asia College, da Chinese University of Hong Kong, da Cornell University, da Lingnan University em Hong Kong e da Helsinki School of Economics and Business Administration. Entre suas publicações há um outro livro sobre economia matemática: *Elements of Dynamic Optimization*, Waveland Press, Inc., 1992. Entre as honrarias que recebeu figuram prêmios da Ford Foundation e auxílios da National Science Foundation, eleição para a presidência da Ohio Association of Economists and Political Scientists, 1963–1964, e citação na publicação *Who's Who in Economics: A Biographical Dictionary of Major Economists 1900–1994*, MIT Press.

Kevin Wainwright é membro do corpo acadêmico do British Columbia Institute of Technology em Burnaby, B.C., Canadá. Desde 2001, ocupa o posto de presidente da associação de docentes e de chefe do programa de Administração de Empresas. Fez sua pós-graduação na Simon Fraser University em Burnaby, B.C., Canadá, e continua a lecionar no Departamento de Economia dessa universidade. Especializou-se em teoria microeconômica e economia matemática.

Prefácio

Este livro foi escrito para estudantes de economia que pretendem aprender os métodos matemáticos básicos que se tornaram indispensáveis para um entendimento adequado da literatura econômica corrente. Infelizmente, para muitos, estudar matemática é assim como tomar um remédio amargo – absolutamente necessário, mas extremamente desagradável. Essa atitude, denominada “medo da matemática”, tem suas raízes – assim acreditamos – em grande parte, na maneira desfavorável com que a matemática geralmente é apresentada aos estudantes. Na crença de que concisão significa elegância, as explicações oferecidas muitas vezes são demasiadamente breves para serem claras, o que deixa os estudantes perplexos e lhes passa um sentimento imerecido de inadequação intelectual. Um estilo de apresentação excessivamente formal, quando não acompanhado por quaisquer ilustrações ou demonstrações intuitivas de “relevância”, pode prejudicar a motivação. Uma progressão desequilibrada do nível do material evidentemente pode fazer com que certos tópicos da matemática pareçam mais difíceis do que realmente são. Finalmente, problemas excessivamente sofisticados apresentados como exercícios tendem a destroçar a confiança dos estudantes em vez de estimular o raciocínio, que é o que se pretende.

Com isso em mente, fizemos um sério esforço para minimizar as características que despertam o medo. Na medida do possível, oferecemos explicações pacientes e não misteriosas. O estilo é deliberadamente informal e “amigável ao leitor”. Como questão de rotina, tentamos prever e responder a perguntas que provavelmente surgirão nas cabeças dos estudantes durante a leitura. Para salientar a relevância da matemática para a economia, deixamos que as necessidades analíticas dos economistas motivassem o estudo das técnicas matemáticas relacionadas e então ilustramos imediatamente essas técnicas com modelos econômicos adequados. Além disso, montamos a caixa de ferramentas matemáticas segundo um esquema cuidadosamente gradativo, no qual as ferramentas elementares servem de base e degrau para as ferramentas mais avançadas discutidas adiante. Sempre que adequado, há representações gráficas que proporcionam um reforço visual aos resultados algébricos. Os problemas e exercícios de fixação foram elaborados como treinamento para ajudar a solidificar a compreensão e a reforçar a confiança, em vez de propor desafios que poderiam inadvertidamente frustrar e intimidar o novato.

Abordamos, neste livro, os seguintes tipos principais de análise econômica: estática (análise de equilíbrio), estática comparativa, problemas de otimização (um tipo especial de estática), dinâmica e otimização dinâmica. Para enfrentar essas análises, são introduzidos os seguintes métodos matemáticos na ocasião oportuna: álgebra matricial, cálculo diferencial e integral, equações diferenciais, equações de diferença e teoria do controle ótimo. Devido à quantidade substancial de modelos econômicos ilustrativos – tanto macro quanto micro – que aparecem neste livro, ele também deverá ser útil para quem já conhece matemática, mas ainda precisa de um guia que o leve do reino da matemática para a terra da economia. Por essa mesma razão, o livro não deve servir apenas como um texto didático para um curso sobre métodos matemáticos, mas também como leitura suplementar em cursos como teoria microeconômica, teoria macroeconômica e crescimento e desenvolvimento econômico.

Tentamos conservar os principais objetivos e o estilo das edições anteriores. Todavia, a presente edição contém diversas modificações significativas. O material sobre programação matemática agora é apresentado mais cedo, em um novo Capítulo 13, intitulado “Tópicos Adicionais de Otimização”. Esse capítulo tem dois temas principais: otimização com limitações de desigualdade e o teorema do envelope. No primeiro tema, as condições de Kuhn-Tucker são desenvolvidas de um modo muito semelhante ao que foi empregado na edição anterior. Contudo, o tópico foi aprimorado com diversas novas aplicações econômicas, entre elas a determinação de preço de carga de pico e o racionamento de consumo. O segundo tema refere-se ao desenvolvimento do teorema do envelope, à função de valor máximo e à noção de dualidade. Aplicando o teorema do envelope a vários modelos econômicos, obtemos resultados importantes como a identidade de Roy, o lema de Shepard e o lema de Hotelling.

A segunda maior adição a esta edição é um novo Capítulo 20 sobre a teoria do controle ótimo. O propósito deste capítulo é apresentar ao leitor o básico do controle ótimo e demonstrar

como ele pode ser aplicado à economia, incluindo exemplos da economia de recursos naturais e da teoria do crescimento ótimo. O material deste capítulo aproveita, em grande parte, a discussão da teoria do controle ótimo em *Elements of Dynamic Optimization*, de autoria de Alpha C. Chiang (McGraw-Hill 1992, agora publicado por Waveland Press, Inc.), que apresenta um tratamento minucioso tanto do controle ótimo quanto de seu precursor, o cálculo de variações.

Além desses dois novos capítulos, há diversos acréscimos e refinamentos significativos nesta edição. No Capítulo 3, ampliamos a discussão da resolução de equações polinomiais de grau mais alto por fatoramento (Seção 3.3). No Capítulo 4, foi acrescentada uma nova seção sobre cadeias de Markov (Seção 4.7). E, no Capítulo 5, introduzimos a verificação da classificação de uma matriz por meio de uma matriz escalonada (Seção 5.1) e a condição de Hawkins-Simon em conexão com o modelo de insumo-produção de Leontief (Seção 5.7). No que diz respeito a aplicações econômicas, foram acrescentados muitos novos exemplos, e algumas das aplicações existentes foram aprimoradas. Foi incluída uma versão linear do modelo IS-LM na Seção 5.6, e uma forma mais geral do modelo na Seção 8.6 foi ampliada para abranger tanto uma economia aberta quanto uma fechada, demonstrando, assim, uma aplicação muito mais rica da estática comparativa a modelos de função geral. Entre outros acréscimos estão uma discussão da utilidade esperada e da preferência de riscos (Seção 9.3), um modelo de maximização de lucros que incorpora a função de produção de Cobb-Douglas (Seção 11.6) e um problema de escolha intertemporal de dois períodos (Seção 12.3). Por fim, os problemas e exercícios de fixação foram revisados e aumentados para dar aos alunos uma oportunidade maior de aprimorar suas habilidades.

Agradecimentos

Há muitas pessoas às quais temos de agradecer pela redação deste livro. Antes de mais nada, devemos muito a todos os matemáticos e economistas cujas idéias originais pavimentam este volume. Em segundo lugar, há muitos alunos cujos esforços e perguntas ao longo dos anos nos ajudaram a moldar a filosofia e a abordagem deste livro.

As três edições anteriores beneficiaram-se dos comentários e sugestões de (em ordem alfabética dos sobrenomes): Nancy S. Barrett, Thomas Birnberg, E. J. R. Booth, Charles E. Butler, Roberta Grower Carey, Emily Chiang, Lloyd R. Cohen, Gary Cornell, Harald Dickson, John C. H. Fei, Warren L. Fisher, Roger N. Folsom, Dennis R. Heffley, Jack Hirshleifer, James C. Hsiao, Ki-Jun Jeong, George Kondor, William F. Lott, Paul B. Manchester, Peter Morgan, Mark Nerlove, J. Frank Sharp, Alan G. Sleeman, Dennis Starleaf, Henry Y. Wan, Jr. e Chiou-Nan Yeh.

Na presente edição, agradecemos sinceramente as sugestões e idéias de Curt L. Anderson, David Andolfatto, James Bathgate, C. R. Birchenhall, Michael Bowe, John Carson, Kimoon Cheong, Youngsub Chun, Kamran M. Dadkhah, Robert Delorme, Patrick Emerson, Roger Nils Folsom, Paul Gomme, Terry Heaps, Suzanne Helburn, Melvin Iyogu, Ki-Jun Jeong, Robbie Jones, John Kane, Heon-Goo Kim, George Kondor, Hui-wen Koo, Stephen Layson, Boon T. Lim, Anthony M. Marino, Richard Miles, Peter Morgan, Rafael Hernández Núñez, Alex Panayides, Xinghe Wang e Hans-Olaf Wieseemann.

Nosso profundo agradecimento a Sarah Dunn, que nos atendeu com tanta capacidade e boa-vontade como digitadora, revisora e assistente de pesquisa. Devemos também um agradecimento especial a Denise Potten por sua dedicação e capacidade logística no estágio de produção. Por fim, estendemos nossa sincera apreciação a Lucille Sutton, Bruce Gin e Lucy Mullins, da McGraw-Hill, por sua paciência e dedicação na produção deste manuscrito. O produto final e quaisquer erros que ainda persistirem são de nossa exclusiva responsabilidade.

Sugestões para o uso deste livro

Devido à acumulação gradativa das ferramentas matemáticas propiciada pela organização deste livro, o modo ideal de estudá-lo é seguir à risca sua seqüência específica de apresentação. Contudo, há algumas alterações possíveis na seqüência de leitura: após concluir as equações diferenciais de primeira ordem (Capítulo 15), o leitor pode passar diretamente para a teoria do controle ótimo (Capítulo 20). Entretanto, se passar diretamente do Capítulo 15 para o Capítulo 20, seria bom que o leitor lesse a Seção 19.5, que trata de diagramas de fase de duas variáveis.

Se a estática comparativa não for uma área de preocupação primordial, o leitor pode omitir a análise estática comparativa de modelos de função geral (Capítulo 8) e passar do Capítulo 7 diretamente para o Capítulo 9. Nesse caso, entretanto, seria necessário omitir também a Seção 11.7, a parte da Seção 12.5 dedicada à estática comparativa, bem como a discussão da dualidade no Capítulo 13.

*Alpha C. Chiang
Kevin Wainwright*

Sumário

PARTE 1

Introdução 1

Capítulo 1

A natureza da economia matemática 3

- 1.1 Economia matemática *versus* economia não-matemática 3
- 1.2 Economia matemática *versus* econometria 4

Capítulo 2

Modelos econômicos 7

- 2.1 Componentes de um modelo matemático 7
 - Variáveis, constantes e parâmetros* 7
 - Equações e identidades* 8
- 2.2 O sistema de números reais 9
- 2.3 O conceito de conjunto 10
 - Notação de conjunto* 10
 - Relação entre conjuntos* 11
 - Operações com conjuntos* 12
 - Leis das operações com conjuntos* 14
 - Exercício 2.3* 15
- 2.4 Relações e funções 16
 - Pares ordenados* 16
 - Relações e funções* 17
 - Exercício 2.4* 20
- 2.5 Tipos de função 20
 - Funções constantes* 20
 - Funções polinomiais* 21
 - Funções racionais* 22
 - Funções não-algébricas* 22
 - Uma digressão sobre expoentes* 22
 - Exercício 2.5* 25
- 2.6 Funções de duas ou mais variáveis independentes 25
- 2.7 Níveis de generalidade 27

PARTE 2

Análise estática (ou de equilíbrio) 29

Capítulo 3

Análise de equilíbrio em economia 31

- 3.1 O significado de equilíbrio 31
- 3.2 Equilíbrio parcial de mercado – um modelo linear 32
 - Construção do modelo* 32
 - Solução por eliminação de variáveis* 33
 - Exercício 3.2* 35
- 3.3 Equilíbrio parcial de mercado – um modelo não-linear 35
 - Equação quadrática versus função quadrática* 36
 - A fórmula quadrática* 37
 - Uma outra solução gráfica* 38
 - Equações polinomiais de grau mais alto* 38
 - Exercício 3.3* 40

3.4	Equilíbrio geral de mercado	41
	<i>Modelo de mercado de duas mercadorias</i>	41
	<i>Exemplo numérico</i>	42
	<i>Caso de n mercadorias</i>	43
	<i>Solução de um sistema geral de equações</i>	44
	<i>Exercício 3.4</i>	45
3.5	Equilíbrio na análise da renda nacional	46
	<i>Exercício 3.5</i>	47

Capítulo 4

Modelos lineares e álgebra matricial 49

4.1	Matrizes e vetores	50
	<i>Matrizes como arranjos</i>	50
	<i>Vetores como matrizes especiais</i>	51
	<i>Exercício 4.1</i>	52
4.2	Operações com matrizes	52
	<i>Adição e subtração de matrizes</i>	52
	<i>Multiplicação escalar</i>	53
	<i>Multiplicação de matrizes</i>	53
	<i>A questão da divisão</i>	57
	<i>A notação Σ</i>	57
	<i>Exercício 4.2</i>	58
4.3	Notas sobre operações vetoriais	59
	<i>Multiplicação de vetores</i>	59
	<i>Interpretação geométrica de operações vetoriais</i>	60
	<i>Dependência linear</i>	62
	<i>Espaço vetorial</i>	63
	<i>Exercício 4.3</i>	65
4.4	Leis comutativas, associativas e distributivas	66
	<i>Adição de matrizes</i>	66
	<i>Multiplicação de matrizes</i>	67
	<i>Exercício 4.4</i>	69
4.5	Matrizes identidades e matrizes nulas	70
	<i>Matrizes identidade</i>	70
	<i>Matrizes nulas</i>	71
	<i>Idiosincrasias da álgebra matricial</i>	71
	<i>Exercício 4.5</i>	72
4.6	Transpostas e inversas	72
	<i>Propriedades das transpostas</i>	73
	<i>Inversas e suas propriedades</i>	74
	<i>Matriz inversa e solução de sistema de equações lineares</i>	76
	<i>Exercício 4.6</i>	77
4.7	Cadeias finitas de Markov	77
	<i>Caso especial: cadeias de Markov absorventes</i>	79
	<i>Exercício 4.7</i>	80

Capítulo 5

Modelos lineares e álgebra matricial (CONTINUAÇÃO) 81

5.1	Condições para a invertibilidade de uma matriz	81
	<i>Condições necessárias versus condições suficientes</i>	81
	<i>Condições para a invertibilidade</i>	83
	<i>Posto de uma matriz</i>	84
	<i>Exercício 5.1</i>	86

5.2	Teste de invertibilidade utilizando determinante	87
	<i>Determinantes e invertibilidade</i>	87
	<i>Cálculo de um determinante de terceira ordem</i>	88
	<i>Cálculo de um determinante de n-ésima ordem por expansão de Laplace</i>	89
	<i>Exercício 5.2</i>	91
5.3	Propriedades básicas de determinantes	92
	<i>Crítério com determinantes para a invertibilidade</i>	94
	<i>Redefinição do posto de uma matriz</i>	95
	<i>Exercício 5.3</i>	96
5.4	Encontrando a matriz inversa	97
	<i>Expansão de um determinante por co-fatores espúrios</i>	97
	<i>Inversão de matriz</i>	98
	<i>Exercício 5.4</i>	100
5.5	Regra de Cramer	101
	<i>Dedução da regra</i>	101
	<i>Observação sobre sistemas homogêneos de equações</i>	103
	<i>Tipos de soluções para um sistema de equações lineares</i>	104
	<i>Exercício 5.5</i>	105
5.6	Aplicação aos modelos de mercado e de renda nacional	105
	<i>Modelo de mercado</i>	105
	<i>Modelo de renda nacional</i>	106
	<i>O modelo IS-LM: economia fechada</i>	107
	<i>Álgebra matricial versus eliminação de variáveis</i>	109
	<i>Exercício 5.6</i>	109
5.7	Modelos de insumo-produto de Leontief	110
	<i>Estrutura de um modelo de insumo-produção</i>	110
	<i>O modelo aberto</i>	111
	<i>Um exemplo numérico</i>	112
	<i>A existência de soluções não negativas</i>	114
	<i>Significado econômico da condição de Hawkins-Simon</i>	116
	<i>O modelo fechado</i>	117
	<i>Exercício 5.7</i>	117
5.8	Limitações da análise estática	118

PARTE 3

Análise estática comparativa 119

Capítulo 6

Estática comparativa e o conceito de derivada 121

6.1	A natureza da estática comparativa	121
6.2	Taxa de variação e a derivada	122
	<i>O quociente de diferenças</i>	122
	<i>A derivada</i>	123
	<i>Exercício 6.2</i>	124
6.3	A derivada e a inclinação de uma curva	124
6.4	O conceito de limite	125
	<i>Limite à esquerda e limite à direita</i>	125
	<i>Ilustrações gráficas</i>	126
	<i>Avaliação de um limite</i>	127
	<i>Visão formal do conceito de limite</i>	128
	<i>Exercício 6.4</i>	131
6.5	Digressão sobre desigualdades e valores absolutos	131
	<i>Regras de desigualdades</i>	131
	<i>Valores absolutos e desigualdades</i>	132

	<i>Solução de uma desigualdade</i>	134
	<i>Exercício 6.5</i>	135
6.6	Teoremas de limite	135
	<i>Teoremas que envolvem uma única função</i>	135
	<i>Teoremas que envolvem duas funções</i>	135
	<i>Limite de uma função polinomial</i>	136
	<i>Exercício 6.6</i>	137
6.7	Continuidade e diferenciabilidade de uma função	137
	<i>Continuidade de uma função</i>	137
	<i>Funções polinomiais e racionais</i>	138
	<i>Diferenciabilidade de uma função</i>	138
	<i>Exercício 6.7</i>	142
 Capítulo 7		
Regras de diferenciação e sua utilização em estática comparativa		143
7.1	Regras de diferenciação para uma função de uma variável	143
	<i>Regra da função constante</i>	143
	<i>Regra da função de potência</i>	144
	<i>Generalização da regra da função de potência</i>	146
	<i>Exercício 7.1</i>	147
7.2	Regras de diferenciação envolvendo duas ou mais funções da mesma variável	147
	<i>Regra da soma/diferença</i>	147
	<i>Regra do produto</i>	149
	<i>Calculando a função de renda marginal a partir da função de renda média</i>	151
	<i>Regra do quociente</i>	153
	<i>Relação entre funções de custo marginal e de custo médio</i>	154
	<i>Exercício 7.2</i>	155
7.3	Regras de diferenciação envolvendo funções de variáveis diferentes	155
	<i>Regra da cadeia</i>	156
	<i>Regra da função inversa</i>	157
	<i>Exercício 7.3</i>	159
7.4	Diferenciação parcial	159
	<i>Derivadas parciais</i>	159
	<i>Técnicas de diferenciação parcial</i>	160
	<i>Interpretação geométrica de derivadas parciais</i>	161
	<i>Vetor gradiente</i>	162
	<i>Exercício 7.4</i>	162
7.5	Aplicações à análise estática comparativa	163
	<i>Modelo de mercado</i>	163
	<i>Modelo de renda nacional</i>	165
	<i>Modelo de insumo-produto</i>	166
	<i>Exercício 7.5</i>	168
7.6	Nota sobre determinantes jacobianos	168
	<i>Exercício 7.6</i>	170

Capítulo 8**Análise estática comparativa de modelos de função geral 171**

8.1	Diferenciais	172
	<i>Diferenciais e derivadas</i>	172
	<i>Diferenciais e elasticidade-ponto</i>	174
	<i>Exercício 8.1</i>	176

8.2	Diferenciais totais	176
	<i>Exercício 8.2</i>	178
8.3	Regras de diferenciais	179
	<i>Exercício 8.3</i>	181
8.4	Derivadas totais	181
	<i>Calculando a derivada total</i>	181
	<i>Uma variação sobre o tema</i>	183
	<i>Mais uma variação sobre o tema</i>	184
	<i>Algumas observações gerais</i>	185
	<i>Exercício 8.4</i>	185
8.5	Derivadas de funções implícitas	185
	<i>Funções implícitas</i>	185
	<i>Derivadas de funções implícitas</i>	187
	<i>Extensão para o caso de equações simultâneas</i>	190
	<i>Exercício 8.5</i>	195
8.6	Estática comparativa de modelos de funções gerais	195
	<i>Modelo de mercado</i>	196
	<i>Abordagem de equações simultâneas</i>	197
	<i>Utilização de derivadas totais</i>	199
	<i>Modelo da renda nacional (IS-LM)</i>	200
	<i>Ampliando o modelo: uma economia aberta</i>	203
	<i>Resumo do procedimento</i>	206
	<i>Exercício 8.6</i>	207
8.7	Limitações da estática comparativa	208

Parte 4

Problemas de otimização 209

Capítulo 9

Otimização: uma variedade especial de análise de equilíbrio 211

9.1	Valores ótimos e valores extremos	211
9.2	Mínimo relativo e máximo relativo: teste da derivada primeira	212
	<i>Extremo relativo versus extremo absoluto</i>	212
	<i>Teste da derivada primeira</i>	213
	<i>Exercício 9.2</i>	216
9.3	Derivadas segundas e derivadas de ordens mais altas	217
	<i>Derivada de uma derivada</i>	217
	<i>Interpretação da derivada segunda</i>	219
	<i>Uma aplicação</i>	221
	<i>Atitudes em relação ao risco</i>	221
	<i>Exercício 9.3</i>	222
9.4	Teste da derivada segunda	223
	<i>Condições necessárias versus condições suficientes</i>	224
	<i>Condições para maximização de lucro</i>	224
	<i>Coefficientes de uma função de custo total cúbica</i>	227
	<i>Curva de receita marginal de inclinação crescente</i>	228
	<i>Exercício 9.4</i>	229
9.5	Série de Maclaurin e série de Taylor	230
	<i>Série de Maclaurin de uma função polinomial</i>	231
	<i>Série de Taylor de uma função polinomial</i>	232
	<i>Expansão de uma função arbitrária</i>	233
	<i>Forma de Lagrange para o resto</i>	236
	<i>Exercício 9.5</i>	237

9.6	Teste da derivada N -ésima para extremo relativo de uma função de uma só variável	238
	<i>Expansão de Taylor e extremo relativo</i>	238
	<i>Alguns casos específicos</i>	238
	<i>Teste da N-ésima derivada</i>	240
	<i>Exercício 9.6</i>	241

Capítulo 10

Funções exponenciais e logarítmicas 243

10.1	A natureza das funções exponenciais	243
	<i>Função exponencial simples</i>	244
	<i>Forma gráfica</i>	244
	<i>Função exponencial generalizada</i>	245
	<i>Uma base preferida</i>	246
	<i>Exercício 10.1</i>	247
10.2	Funções exponenciais naturais e o problema do crescimento	248
	<i>O número e</i>	248
	<i>Uma interpretação econômica de e</i>	249
	<i>Juro composto e a função Ae^{rt}</i>	250
	<i>Taxa instantânea de crescimento</i>	251
	<i>Crescimento contínuo versus crescimento discreto</i>	252
	<i>Desconto e crescimento negativo</i>	253
	<i>Exercício 10.2</i>	254
10.3	Logaritmos	254
	<i>O significado de logaritmo</i>	254
	<i>Log comum e Log natural</i>	255
	<i>Regras de logaritmos</i>	256
	<i>Uma aplicação</i>	258
	<i>Exercício 10.3</i>	258
10.4	Funções logarítmicas	259
	<i>Funções log e funções exponenciais</i>	259
	<i>A forma gráfica</i>	259
	<i>Conversão de base</i>	261
	<i>Exercício 10.4</i>	262
10.5	Derivadas de funções exponenciais e logarítmicas	263
	<i>Regra da função logarítmica</i>	263
	<i>Regra da função exponencial</i>	264
	<i>As regras generalizadas</i>	264
	<i>O caso da base b</i>	266
	<i>Derivadas de ordens mais altas</i>	266
	<i>Uma aplicação</i>	267
	<i>Exercício 10.5</i>	268
10.6	Tempo ótimo	268
	<i>Um problema de armazenagem de vinho</i>	268
	<i>Condições de maximização</i>	269
	<i>Um problema de corte de madeira</i>	271
	<i>Exercício 10.6</i>	272
10.7	Outras aplicações de derivadas exponenciais e logarítmicas	272
	<i>Achando a taxa de crescimento</i>	272
	<i>Taxa de crescimento de uma combinação de funções</i>	273
	<i>Achando a elasticidade pontual</i>	274
	<i>Exercício 10.7</i>	275

Capítulo 11**O caso de mais de uma variável de escolha 277**

- 11.1** A versão diferencial de condições de otimização 277
 Condição de primeira ordem 277
 Condição de segunda ordem 278
 Condições de diferencial versus condições de derivada 279
- 11.2** Valores extremos de uma função de duas variáveis 279
 Condição de primeira ordem 279
 Derivadas parciais de segunda ordem 281
 Diferencial total de segunda ordem 282
 Condição de segunda ordem 283
 Exercício 11.2 286
- 11.3** Formas quadráticas – uma excursão 286
 Diferencial total de segunda ordem como uma forma quadrática 286
 Formas negativas definidas e positivas definidas 287
 Teste com determinantes para condição de sinal definido 287
 Formas quadráticas com três variáveis 290
 Formas quadráticas de n variáveis 292
 Teste da raiz característica para condição de sinal definido 292
 Exercício 11.3 296
- 11.4** Funções objetivo com mais de duas variáveis 297
 Condição de primeira ordem para extremo 297
 Condição de segunda ordem 298
 O caso de n variáveis 301
 Exercício 11.4 302
- 11.5** Condições de segunda ordem relativas à concavidade e convexidade 302
 Verificação de concavidade e convexidade 304
 Funções diferenciáveis 308
 Funções convexas versus conjuntos convexos 310
 Exercício 11.5 314
- 11.6** Aplicações na economia 314
 Problema de uma empresa com vários produtos 314
 Discriminação de preços 317
 Decisões de insumos de uma empresa 319
 Exercício 11.6 324
- 11.7** Aspectos de estática comparativa da otimização 325
 Soluções de forma reduzida 325
 Modelos de função geral 325
 Exercício 11.7 328

Capítulo 12**Otimização com restrições de igualdade 329**

- 12.1** Efeitos de uma restrição 329
- 12.2** Achando os valores estacionários 331
 Método do multiplicador de Lagrange 331
 Abordagem da diferencial total 333
 Uma interpretação do multiplicador de Lagrange 334
 Casos de n variáveis e múltiplas restrições 336
 Exercício 12.2 337
- 12.3** Condições de segunda ordem 337
 Diferencial total de segunda ordem 338
 Condições de segunda ordem 339
 O hessiano aumentado 339
 Caso de n variáveis 342

	<i>Caso de múltiplas restrições</i>	344
	<i>Exercício 12.3</i>	344
12.4	Quase-concavidade e quase-convexidade	345
	<i>Caracterização geométrica</i>	345
	<i>Definição algébrica</i>	347
	<i>Funções diferenciáveis</i>	349
	<i>Um exame mais aprofundado do hessiano aumentado</i>	352
	<i>Extremos absolutos versus extremos relativos</i>	353
	<i>Exercício 12.4</i>	354
12.5	Maximização da utilidade e demanda do consumidor	355
	<i>Condição de primeira ordem</i>	355
	<i>Condição de segunda ordem</i>	356
	<i>Análise estática comparativa</i>	358
	<i>Variações proporcionais em preços e renda</i>	362
	<i>Exercício 12.5</i>	362
12.6	Funções homogêneas	363
	<i>Homogeneidade linear</i>	364
	<i>A função produção de Cobb-Douglas</i>	366
	<i>Extensões dos resultados</i>	368
	<i>Exercício 12.6</i>	369
12.7	Combinação de insumos de custo mínimo	370
	<i>Condição de primeira ordem</i>	370
	<i>Condição de segunda ordem</i>	371
	<i>A rota de expansão</i>	372
	<i>Funções homotéticas</i>	373
	<i>Elasticidade de substituição</i>	375
	<i>Função de produção CES</i>	376
	<i>Função de Cobb-Douglas como um caso especial da função CES</i>	378
	<i>Exercício 12.7</i>	380
 Capítulo 13		
Tópicos adicionais de otimização		381
13.1	Programação não-linear e condições de Kuhn-Tucker	381
	<i>Etapa 1: Efeito de restrições de não-negatividade</i>	382
	<i>Etapa 2: Efeito de restrições de desigualdade</i>	383
	<i>Interpretação das condições de Kuhn-Tucker</i>	387
	<i>O caso de n variáveis, m restrições</i>	388
	<i>Exercício 13.1</i>	390
13.2	A qualificação da restrição	391
	<i>Irregularidades nos pontos de fronteira</i>	391
	<i>A qualificação de restrição</i>	393
	<i>Restrições lineares</i>	395
	<i>Exercício 13.2</i>	396
13.3	Aplicações econômicas	397
	<i>Racionamento em tempo de guerra</i>	397
	<i>Determinação de preço de pico (carga máxima)</i>	399
	<i>Exercício 13.3</i>	402
13.4	Teoremas de suficiência em programação não-linear	403
	<i>O teorema de suficiência de Kuhn-Tucker: programação côncava</i>	403
	<i>O teorema de suficiência de Arrow-Enthoven: programação quase-côncava</i>	404
	<i>Um teste de qualificação de restrição</i>	405
	<i>Exercício 13.4</i>	405
13.5	Funções de valor máximo e o teorema do envelope	406
	<i>O teorema do envelope para otimização sem restrição</i>	406

	<i>A função lucro</i>	407
	<i>Condição de reciprocidade</i>	408
	<i>O teorema do envelope para otimização com restrição</i>	411
	<i>Interpretação do multiplicador de Lagrange</i>	412
13.6	Dualidade e o teorema do envelope	413
	<i>O problema primordial</i>	413
	<i>O problema dual</i>	414
	<i>Dualidade</i>	415
	<i>Identidade de Roy</i>	415
	<i>O lema de Shephard</i>	416
	<i>Exercício 13.6</i>	420
13.7	Algumas observações finais	421

PARTE 5

Análise dinâmica 423

Capítulo 14

Economia dinâmica e cálculo integral 425

14.1	Dinâmica e integração	425
14.2	Integrais indefinidas	427
	<i>A natureza das integrais</i>	427
	<i>Regras básicas de integração</i>	427
	<i>Regras de operação</i>	429
	<i>Regras que envolvem substituição</i>	432
	<i>Exercício 14.2</i>	434
14.3	Integrais definidas	435
	<i>Significado de integrais definidas</i>	435
	<i>Uma integral definida como uma área sob uma curva</i>	436
	<i>Algumas propriedades de integrais definidas</i>	439
	<i>Um outro modo de ver a integral indefinida</i>	440
	<i>Exercício 14.3</i>	440
14.4	Integrais impróprias	441
	<i>Limites de integração infinitos</i>	441
	<i>Integrando infinito</i>	443
	<i>Exercício 14.4</i>	444
14.5	Algumas aplicações econômicas de integrais	444
	<i>De uma função marginal para uma função total</i>	444
	<i>Investimento e formação de capital</i>	445
	<i>Valor presente de um fluxo de caixa</i>	447
	<i>Valor presente de um fluxo perpétuo</i>	449
	<i>Exercício 14.5</i>	450
14.6	Modelo de crescimento de Domar	450
	<i>A estrutura</i>	450
	<i>Encontrando a solução</i>	451
	<i>O fio da navalha</i>	452
	<i>Exercício 14.6</i>	453

Capítulo 15

Tempo contínuo: equações diferenciais de primeira ordem 455

15.1	Equações diferenciais lineares de primeira ordem com coeficiente constante e termo constante	455
	<i>O caso homogêneo</i>	456
	<i>O caso não-homogêneo</i>	456

	<i>Verificação da solução</i>	458
	<i>Exercício 15.1</i>	459
15.2	Dinâmica do preço de mercado	459
	<i>A estrutura</i>	459
	<i>A trajetória temporal</i>	460
	<i>A estabilidade dinâmica de equilíbrio</i>	460
	<i>Uma utilização alternativa do modelo</i>	461
	<i>Exercício 15.2</i>	462
15.3	Coeficiente variável e termo variável	463
	<i>O caso homogêneo</i>	463
	<i>O caso não-homogêneo</i>	464
	<i>Exercício 15.3</i>	465
15.4	Equações diferenciais exatas	466
	<i>Equações diferenciais exatas</i>	466
	<i>Método de solução</i>	467
	<i>Fator de integrante</i>	469
	<i>Solução de equações diferenciais lineares de primeira ordem</i>	469
	<i>Exercício 15.4</i>	471
15.5	Equações diferenciais não-lineares de primeira ordem e de primeiro grau	471
	<i>Equações diferenciais exatas</i>	471
	<i>Variáveis separáveis</i>	472
	<i>Equações redutíveis à forma linear</i>	473
	<i>Exercício 15.5</i>	474
15.6	A abordagem gráfico-qualitativa	474
	<i>O diagrama de fase</i>	475
	<i>Tipos de trajetória temporal</i>	475
	<i>Exercício 15.6</i>	477
15.7	Modelo de crescimento de Solow	477
	<i>A estrutura</i>	478
	<i>Uma análise gráfico-qualitativa</i>	479
	<i>Uma ilustração quantitativa</i>	480
	<i>Exercício 15.7</i>	481
Capítulo 16		
Equações diferenciais de ordem mais alta 483		
16.1	Equações diferenciais lineares de segunda ordem com coeficientes constantes e termo constante	484
	<i>A solução particular</i>	484
	<i>A função complementar</i>	485
	<i>A estabilidade dinâmica de equilíbrio</i>	489
	<i>Exercício 16.1</i>	490
16.2	Números complexos e funções circulares	491
	<i>Números imaginários e complexos</i>	491
	<i>Raízes complexas</i>	492
	<i>Funções circulares</i>	493
	<i>Propriedades das funções seno e co-seno</i>	494
	<i>Relações de Euler</i>	496
	<i>Representações alternativas de números complexos</i>	498
	<i>Exercício 16.2</i>	500
16.3	Análise do caso da raiz complexa	501
	<i>A função complementar</i>	501
	<i>Um exemplo de solução</i>	502
	<i>A trajetória temporal</i>	504
	<i>A estabilidade dinâmica do equilíbrio</i>	505
	<i>Exercício 16.3</i>	506

16.4	Um modelo de mercado com expectativas de preço	506
	<i>Tendência de preço e expectativas de preço</i>	506
	<i>Um modelo simplificado</i>	507
	<i>A trajetória temporal do preço</i>	507
	<i>Exercício 16.4</i>	510
16.5	A interação entre inflação e desemprego	511
	<i>A relação de Phillips</i>	511
	<i>A relação de Phillips com expectativas aumentadas</i>	511
	<i>O retorno da inflação para o desemprego</i>	512
	<i>A trajetória temporal do π</i>	513
	<i>Exercício 16.5</i>	515
16.6	Equações diferenciais com um termo variável	516
	<i>Método de coeficientes indeterminados</i>	516
	<i>Uma modificação</i>	517
	<i>Exercício 16.6</i>	518
16.7	Equações diferenciais lineares de ordem mais alta	518
	<i>Encontrando a solução</i>	518
	<i>Convergência e o teorema de Routh</i>	520
	<i>Exercício 16.7</i>	521
 Capítulo 17		
Tempo discreto: equações de diferenças de primeira ordem		523
17.1	Tempo discreto, diferenças e equações de diferenças	523
17.2	Resolvendo uma equação de diferenças de primeira ordem	525
	<i>Método iterativo</i>	525
	<i>Método geral</i>	527
	<i>Exercício 17.2</i>	529
17.3	A estabilidade dinâmica de equilíbrio	530
	<i>O significado de b</i>	530
	<i>O papel de A</i>	531
	<i>Convergência ao equilíbrio</i>	531
	<i>Exercício 17.3</i>	533
17.4	O modelo da teia de aranha	533
	<i>O modelo</i>	533
	<i>As teias de aranha</i>	534
	<i>Exercício 17.4</i>	536
17.5	Um modelo de mercado com estoque	537
	<i>O modelo</i>	537
	<i>A trajetória temporal</i>	538
	<i>Resumo gráfico dos resultados</i>	539
	<i>Exercício 17.5</i>	540
17.6	Equação de diferenças não-linear – a abordagem gráfico-qualitativa	540
	<i>Diagrama de fase</i>	540
	<i>Tipos de trajetória temporal</i>	542
	<i>Um mercado com um teto de preço</i>	543
	<i>Exercício 17.6</i>	544

Capítulo 18**Equações de diferenças de ordens mais altas** 545

18.1	Equações de diferenças lineares de segunda ordem com coeficientes constantes e termo constante	546
	<i>Solução particular</i>	546
	<i>Função complementar</i>	547
	<i>A convergência da trajetória temporal</i>	550
	<i>Exercício 18.1</i>	552

18.2	Modelo de Samuelson para a interação multiplicador-aceleração	552
	<i>A estrutura</i>	553
	<i>A solução</i>	553
	<i>Convergência versus divergência</i>	554
	<i>Um resumo gráfico</i>	556
	<i>Exercício 18.2</i>	557
18.3	Inflação e desemprego em tempo discreto	558
	<i>O modelo</i>	558
	<i>A equação de diferenças em p</i>	558
	<i>A trajetória temporal de p</i>	559
	<i>A análise de U</i>	560
	<i>A relação de Phillips de longo prazo</i>	561
	<i>Exercício 18.3</i>	562
18.4	Generalizações para equações de termo variável e ordens mais altas	562
	<i>Termo variável na forma de cm^i</i>	562
	<i>Termo variável na forma de ct^n</i>	564
	<i>Equações de diferenças lineares de ordens mais altas</i>	565
	<i>Convergência e o teorema de Schur</i>	566
	<i>Exercício 18.4</i>	568
 Capítulo 19		
Equações diferenciais e equações de diferenças simultâneas		569
19.1	A gênese de sistemas dinâmicos	569
	<i>Padrões de variação interativos</i>	569
	<i>A transformação de uma equação dinâmica de ordem alta</i>	570
19.2	Resolvendo equações dinâmicas simultâneas	571
	<i>Equações de diferenças simultâneas</i>	571
	<i>Notação matricial</i>	573
	<i>Equações diferenciais simultâneas</i>	575
	<i>Comentários adicionais sobre a equação característica</i>	578
	<i>Exercício 19.2</i>	579
19.3	Modelos dinâmicos de insumo-produto	580
	<i>Defasagem de tempo na produção</i>	580
	<i>Excesso de demanda e ajuste da produção</i>	582
	<i>Formação de capital</i>	583
	<i>Exercício 19.3</i>	585
19.4	Mais uma vez, o modelo inflação-desemprego	586
	<i>Equações diferenciais simultâneas</i>	586
	<i>Trajетórias de solução</i>	586
	<i>Equações de diferenças simultâneas</i>	589
	<i>Trajетórias de solução</i>	589
	<i>Exercício 19.4</i>	590
19.5	Diagramas de fase de duas variáveis	590
	<i>O espaço de fase</i>	591
	<i>As curvas de demarcação</i>	591
	<i>Linhas de fluxo</i>	593
	<i>Tipos de equilíbrio</i>	594
	<i>Inflação e regra monetária à moda de Obst</i>	595
	<i>Exercício 19.5</i>	598
19.6	Linearização de um sistema de equações diferenciais não-linear	599
	<i>Expansão de Taylor e linearização</i>	599
	<i>A linearização homogênea</i>	600
	<i>Análise da estabilidade local</i>	601
	<i>Exercício 19.6</i>	605

**Capítulo 20****Teoria do controle ótimo 607**

- 20.1** A natureza do controle ótimo 607
Ilustração: um modelo macroeconômico simples 608
Princípio do máximo de Pontryagin 609
- 20.2** Condições terminais alternativas 614
Ponto terminal fixo 614
Reta terminal horizontal 615
Reta terminal vertical truncada 615
Reta terminal horizontal truncada 615
Exercício 20.2 618
- 20.3** Problemas autônomos 619
- 20.4** Aplicações econômicas 620
Maximização do tempo de vida da utilidade 620
Recurso exaurível 621
Exercício 20.4 623
- 20.5** Horizonte de tempo infinito 624
Modelo neoclássico de crescimento ótimo 624
A hamiltoniana de valor corrente 626
Construindo um diagrama de fase 626
Análise do diagrama de fase 627
- 20.6** Limitações da análise dinâmica 628

O alfabeto grego 631**Símbolos matemáticos 631****Bibliografia recomendada 635****Respostas a exercícios selecionados 637****Índice 651**

PARTE 1

Introdução



CAPÍTULO 1

A natureza da economia matemática

A **economia matemática** não é um ramo distinto da economia no mesmo sentido das finanças públicas e do comércio exterior. Trata-se mais de uma *abordagem* da análise econômica na qual o economista utiliza símbolos matemáticos para enunciar um problema, recorrendo também a teoremas matemáticos conhecidos para auxiliar o raciocínio. Quanto ao tópico específico da análise, pode ser teoria microeconômica ou macroeconômica, finanças públicas, economia urbana ou outro tema qualquer.

Utilizando o termo *economia matemática* no sentido mais amplo possível, pode-se perfeitamente afirmar que, hoje, todo livro didático elementar de economia menciona exemplos de economia matemática, considerando-se que métodos geométricos são freqüentemente utilizados para obter resultados teóricos. Mais comumente, entretanto, a economia matemática é reservada para descrever casos empregando técnicas matemáticas mais avançadas que a simples geometria, tais como álgebra matricial, cálculo diferencial e integral, equações diferenciais etc. O propósito deste livro é apresentar ao leitor os aspectos mais fundamentais desses métodos matemáticos – os que se encontram diariamente na literatura econômica atual.

1.1 Economia matemática *versus* economia não-matemática

Visto que a economia matemática é meramente uma abordagem da análise econômica, não deve ser, e não é, fundamentalmente diferente da abordagem *não*-matemática da análise econômica. O propósito de qualquer análise teórica, independentemente da abordagem, é sempre derivar um conjunto de conclusões ou teoremas a partir de um dado conjunto de premissas ou postulados via um processo de raciocínio. Há duas diferenças principais entre “economia matemática” e “economia literária”: a primeira é que, na economia matemática, as premissas e conclusões são enunciadas em símbolos matemáticos e não em palavras, e em equações em vez de sentenças. A segunda é que, em lugar da lógica literária, são usados teoremas matemáticos – há uma profusão deles a nosso dispor – no processo de raciocínio. Considerando que símbolos e palavras são realmente equivalentes (confirmado pelo fato de que símbolos usualmente são definidos em palavras), pouco importa qual deles é escolhido. Porém, provavelmente ninguém contestaria que símbolos são mais convenientes para usar no raciocínio dedutivo e que, certamente, são mais favoráveis à concisão e à precisão de enunciados.

A escolha entre lógica literária e lógica matemática é, mais uma vez, uma questão de pouca importância, mas a matemática tem a vantagem de obrigar os analistas a enunciar suas premissas explicitamente em cada estágio do raciocínio. Isso porque teoremas matemáticos usualmente são enunciados na forma “se-então”, de modo que, para poder usar a parte do “então” (o resulta-

do) do teorema, os analistas primeiramente têm de ter certeza de que a parte do “se” (condição) está de acordo com as premissas explícitas adotadas.

Mesmo admitindo esses pontos, ainda pode-se perguntar por que é necessário ir além dos métodos geométricos. A resposta é que, conquanto a análise geométrica tenha a importante vantagem de ser visual, também está sujeita a sérias limitações dimensionais. Na discussão geométrica normal das curvas de indiferença, por exemplo, a premissa padrão considera que há somente *duas* mercadorias disponíveis para o consumidor. Essa premissa simplificadora não é adotada voluntariamente, mas imposta, porque desenhar um gráfico tridimensional é extremamente difícil e construir um gráfico de quatro dimensões (ou mais) é, na realidade, uma impossibilidade física. Para tratar do caso mais geral de 3, 4 ou n mercadorias, devemos recorrer a uma ferramenta mais flexível: as equações. Essa única razão já seria motivação suficiente para o estudo de métodos matemáticos além da geometria.

Em suma, vemos que a abordagem matemática teria razão de proclamar as seguintes vantagens: (1) a “linguagem” utilizada é mais concisa e precisa; (2) existe uma profusão de teoremas matemáticos a nosso dispor; (3) como nos obriga a enunciar explicitamente todas as nossas premissas como um pré-requisito da utilização dos teoremas matemáticos, ela nos resguarda da armadilha de adotar inadvertidamente premissas implícitas indesejáveis; e (4) nos permite tratar o caso geral de n variáveis.

Contrapondo-se a essas vantagens, às vezes ouvimos críticas afirmando que uma teoria obtida matematicamente é inevitavelmente *não-realista*. Entretanto, essa crítica não é válida. Na verdade, o epíteto “não-realista” não pode ser utilizado nem mesmo para criticar a teoria econômica em geral, quer a abordagem seja ou não matemática. Teoria é, por sua própria natureza, uma abstração do mundo real. É um recurso para isolar apenas os fatores e as relações mais essenciais, de modo que possamos estudar o ponto crucial do problema que temos em mãos, livre das inúmeras complicações que existem no mundo real. Assim, a afirmação “falta realismo à teoria” é meramente um truísmo que não pode ser aceito como uma crítica válida da teoria. Pelo mesmo critério, não tem muito sentido designar qualquer uma das abordagens da teoria como “não-realista”. Por exemplo, a teoria da empresa sob concorrência perfeita não é realista, assim como a teoria da empresa sob concorrência imperfeita; porém, se essas teorias são obtidas matematicamente ou não é irrelevante e imaterial.

Para aproveitar as vantagens da profusão de ferramentas matemáticas, é claro que, antes de mais nada, devemos adquirir essas ferramentas. Infelizmente, as ferramentas de interesse para o economista estão amplamente espalhadas por muitos cursos de matemática – um número demasiadamente grande para caber razoavelmente no plano de estudos de um estudante normal de economia. O serviço prestado por este volume é reunir em um só lugar os métodos matemáticos mais relevantes para a literatura econômica, organizá-los em uma ordem de progressão lógica, explicar completamente cada método e, então, ilustrar imediatamente como ele é aplicado em análise econômica. Vinculando os métodos e suas aplicações, a relevância da matemática para a economia fica mais transparente do que seria possível em cursos normais de matemática nos quais as aplicações ilustradas são predominantemente vinculadas à física e à engenharia. Por conseguinte, a familiaridade com o conteúdo deste livro (e, se possível, também com seu volume subsequente: Alpha C. Chiang: *Elements of Dynamic Optimization*, McGraw-Hill, 1992, publicado agora pela Waveland Press, Inc.) deve habilitar o leitor a compreender a maioria dos artigos escritos por profissionais encontrados em periódicos como o *American Economic Review*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Review of Economics and Statistics* e *Economic Journal*. Os leitores que, ao tomarem conhecimento da matemática econômica, desenvolverem um sério interesse pelo assunto poderão partir para um estudo mais rigoroso e avançado da matemática.

1.2 Economia matemática versus econometria

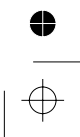
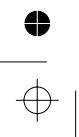
O termo *economia matemática* às vezes é confundido com um termo relacionado: *econometria*. Como a parte “metria” do último termo dá a entender, a econometria preocupa-se principalmente com a medição de dados econômicos. Por conseguinte, trata do estudo de observações em-

píricas utilizando métodos estatísticos de estimação e teste de hipóteses. A economia matemática, por outro lado, refere-se à aplicação da matemática aos aspectos puramente *teóricos* da análise econômica, preocupando-se muito pouco, ou quase nada, com problemas estatísticos como erros de medição das variáveis que estão sendo estudadas.

Neste volume, nos limitaremos à economia matemática. Isto é, nos concentraremos na aplicação da matemática ao raciocínio dedutivo e não ao estudo indutivo e, conseqüentemente, estaremos lidando primordialmente com material teórico e não empírico. É claro que isso é uma questão exclusivamente de escolha do escopo da discussão e não quer dizer, de modo algum, que a econometria é menos importante.

Na verdade, estudos empíricos e análises teóricas freqüentemente são complementares e se reforçam mutuamente. Por um lado, a validade das teorias tem de ser testada em relação a dados empíricos antes que elas possam ser aplicadas com confiança. Por outro lado, o trabalho estatístico necessita da teoria econômica como guia, para determinar a direção mais relevante e proveitosa da pesquisa.

Entretanto, de um certo modo, a economia matemática pode ser considerada a mais básica das duas, pois, para realizar um estudo estatístico e econométrico significativo, é indispensável uma boa estrutura teórica – de preferência em formulação matemática. Portanto, o assunto deste volume deverá ser útil não somente para os interessados na economia teórica, mas também para os que buscam uma base para prosseguir em estudos econométricos.



CAPÍTULO 2

Modelos econômicos

Como mencionamos anteriormente, qualquer teoria econômica é, necessariamente, uma abstração do mundo real. Uma razão é que a imensa complexidade da economia real faz com que seja impossível entender todas as inter-relações de uma vez só e, a propósito, nem todas essas inter-relações têm a mesma importância para o entendimento do fenômeno econômico específico em estudo. Portanto, o procedimento sensato é escolher aquilo que a nossa razão determinar como fatores primários e relações relevantes para o nosso problema e concentrar nossa atenção somente neles. Essa estrutura analítica deliberadamente simplificada é denominada um *modelo econômico*, já que é apenas uma representação esquemática e aproximada da economia real.

2.1 Componentes de um modelo matemático

Um modelo econômico é apenas uma estrutura teórica e não há nenhuma razão inerente por que deva ser matemático. Contudo, se o modelo *for* matemático, usualmente consistirá em um conjunto de *equações* elaborado para descrever a estrutura do modelo. Relacionando algumas *variáveis* entre si de certas maneiras, essas equações dão forma matemática ao conjunto de premissas analíticas adotadas. Então, por meio da aplicação de operações matemáticas relevantes a essas equações, podemos tentar obter um conjunto de conclusões que resultem logicamente dessas premissas.

Variáveis, constantes e parâmetros

Uma *variável* é algo cujo valor pode mudar, isto é, algo que pode assumir valores diferentes. Entre as variáveis freqüentemente usadas na economia estão preço, lucro, receita, custo, renda nacional, consumo, investimento, importações e exportações. Visto que cada variável pode assumir diversos valores, ela deve ser representada por um símbolo e não por um número específico. Por exemplo, podemos representar preço por P , lucro por π , receita por R , custo por C , renda nacional por Y e assim por diante. Entretanto, quando escrevemos $P = 3$ ou $C = 18$, estamos “congelando” essas variáveis em valores específicos (em unidades adequadamente escolhidas).

Quando construído de maneira apropriada, um modelo econômico pode ser resolvido e nos dar *valores de solução* de um determinado conjunto de variáveis, tais como o nível de preço de equilíbrio do mercado ou o nível de produção para a maximização dos lucros. Essas variáveis cujos valores de solução procuramos utilizando o modelo são conhecidas como *variáveis endógenas* (que se originam de dentro). Contudo, o modelo também pode conter variáveis que por supor-

mos ser determinadas por forças externas ao modelo e cujos valores são aceitos somente como dados; essas variáveis são denominadas *variáveis exógenas* (que se originam de fora). Deve-se notar que uma variável endógena em um modelo pode perfeitamente ser exógena em outro. No caso de uma análise da determinação do preço de mercado do trigo, (P), por exemplo, a variável P deve ser definitivamente endógena; mas, na estrutura de uma teoria de gastos de consumidor, P se tornaria um dado do consumidor individual e, portanto, deve ser considerada exógena.

Variáveis freqüentemente aparecem combinadas com números fixos ou constantes, tal como nas expressões $7P$ ou $0,5R$. Uma *constante* é uma grandeza que não muda e, portanto, é a antítese de uma variável. Quando uma constante está ligada a uma variável, ela é normalmente denominada o *coeficiente* daquela variável. Contudo, um coeficiente pode ser simbólico em vez de numérico. Podemos, por exemplo, usar o símbolo a para representar uma dada constante e utilizar em um modelo a expressão aP em lugar de $7P$ para obter um grau mais alto de generalidade (ver Seção 2.7). O símbolo a é um caso bastante peculiar – deve representar uma dada constante mas, mesmo assim, como não lhe designamos um número específico, ele pode assumir praticamente qualquer valor. Em suma, é uma *constante* que é *variável*! Para identificar seu status especial, damos-lhe o nome distintivo de *constante paramétrica* (ou, simplesmente, *parâmetro*).

Devemos enfatizar também que, embora valores diferentes possam ser atribuídos a um parâmetro, ele, não obstante, deve ser considerado como um dado no modelo. É por essa razão que dizemos simplesmente “constante”, mesmo quando a constante é paramétrica. Nesse aspecto particular, parâmetros são muito semelhantes a variáveis exógenas, pois ambos devem ser tratados como “dados” em um modelo. Isso explica por que muitos autores, por simplicidade, referem-se a ambos, coletivamente, pela simples designação “parâmetros”.

Por questão de convenção, constantes paramétricas normalmente são representadas pelos símbolos a , b , c ou por seus correspondentes no alfabeto grego: α , β e γ . Mas é claro que outros símbolos também são permitidos. Quanto às variáveis exógenas, para que elas possam ser distinguidas visualmente de suas primas endógenas, adotaremos a prática de ligar um subscrito 0 ao símbolo escolhido. Por exemplo, se P simbolizar preço, então P_0 significa um preço determinado exogenamente.

Equações e identidades

Variáveis podem existir independentemente, mas só se tornam realmente interessantes quando estão relacionadas umas com as outras por equações ou por desigualdades. Por enquanto, vamos discutir apenas equações.

Podemos distinguir três tipos de equações em aplicações econômicas: equações definicionais, equações comportamentais e equações condicionais.

Uma *equação definicional* estabelece uma identidade entre duas expressões alternativas que têm exatamente o mesmo significado. Nesse tipo de equação, o sinal de identidade $\equiv$ (lê-se: “é idêntico a”) é freqüentemente empregado no lugar do sinal de igual normal $=$, embora este último também seja aceitável. Por exemplo, o lucro total é definido como o resultado da receita total menos o custo total; portanto, podemos escrever

$$\pi \equiv R - C$$

Uma *equação comportamental*, por outro lado, especifica o modo como uma variável se comporta em resposta a mudanças em outras variáveis. Isso pode envolver comportamento humano (como o padrão de consumo agregado em relação à renda nacional) ou comportamento não-humano (por exemplo, o modo como uma empresa reage a mudanças nos valores de produção). Segundo uma definição geral, equações comportamentais podem ser utilizadas para descrever o contexto institucional geral de um modelo, incluindo aspectos tecnológicos (por exemplo, a função de produção) e legais (por exemplo, a estrutura tributária). Contudo, antes de poder escrever uma equação comportamental, é sempre necessário adotar premissas definidas a respeito do padrão comportamental da variável em questão. Considere as duas funções de custo

$$C = 75 + 10Q \quad (2.1)$$

$$C = 110 + Q^2 \quad (2.2)$$

onde Q denota a quantidade de produção. Visto que as formas das duas equações são diferentes, a condição de produção admitida em cada uma é obviamente diferente da outra. Em (2.1), o custo fixo (o valor de C quando $Q=0$) é 75, ao passo que em (2.2) é 110. A variação do custo também é diferente. Em (2.1), para cada unidade de aumento em Q , há um aumento constante de 10 em C . Mas, em (2.2), à medida que Q aumenta, unidade após unidade, C aumentará de quantidades progressivamente maiores. É claro que é por meio da especificação da forma das equações comportamentais que damos expressão matemática às premissas adotadas para um modelo.

O terceiro tipo, a *equação condicional*, determina um requisito a ser satisfeito. Por exemplo, em um modelo que envolve a noção de equilíbrio, devemos estabelecer uma *condição de equilíbrio* que descreva o pré-requisito para atingir o equilíbrio. Duas das condições de equilíbrio mais familiares em economia são

$$Q_d = Q_s \quad [\text{quantidade demandada} = \text{quantidade fornecida}]$$

e $S = I \quad [\text{poupança pretendida} = \text{investimento pretendido}]$

pertinentes, respectivamente, ao equilíbrio de um modelo de mercado e ao equilíbrio da renda nacional em sua forma mais simples. De modo semelhante, um modelo de otimização ou obtém ou aplica uma ou mais *condições de otimização*. Uma dessas condições que vêm à mente com muita facilidade é a condição

$$MC = MR \quad [\text{custo marginal} = \text{receita marginal}]$$

na teoria da empresa. Como as equações desse tipo não são nem definicionais nem comportamentais, constituem uma classe por si próprias.

2.2 O sistema de números reais

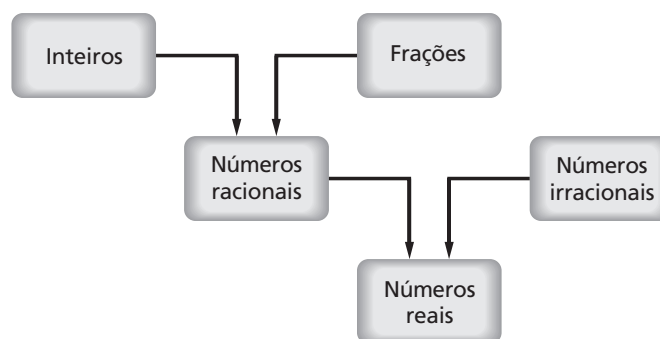
Equações e variáveis são os componentes essenciais de um modelo matemático. Mas, visto que os valores que uma variável econômica assume usualmente são numéricos, seria bom falar um pouco sobre o sistema numérico. Aqui, trataremos apenas dos denominados números reais.

Números inteiros como 1, 2, 3... são denominados *inteiros positivos*; são os números mais frequentemente utilizados em contagem. Suas contrapartes negativas $-1, -2, -3...$ são denominadas *inteiros negativos*; podem ser empregados, por exemplo, para indicar temperaturas abaixo de zero (em graus). O número zero (0), por outro lado, não é nem positivo nem negativo e, nesse sentido, é único. Vamos reunir todos os inteiros positivos e negativos e o número zero em uma única categoria, denominando-os, coletivamente, *conjunto de todos os inteiros*.

É claro que números inteiros não esgotam todos os números possíveis, pois temos *frações*, como $\frac{2}{3}, \frac{5}{4}$ e $\frac{7}{3}$, que – se colocadas sobre uma régua – cairiam entre os inteiros. Temos também frações negativas, como $-\frac{1}{2}$ e $-\frac{2}{5}$. Juntos, esses números formam o *conjunto de todas as frações*.

A propriedade comum a todos os números fracionários é que cada um deles pode ser expresso como uma razão entre dois inteiros. Qualquer número que possa ser expresso como uma razão entre dois inteiros é denominado um *número racional*. Mas os próprios inteiros também são racionais, porque qualquer inteiro n pode ser considerado como a razão $n/1$. Juntos, o conjunto de todos os inteiros e o conjunto de todas as frações formam o *conjunto de todos os números racionais*. Uma característica alternativa que define um número racional é que ele pode ser expresso como uma fração decimal exata (por exemplo, $\frac{1}{4} = 0,25$) ou como uma fração decimal periódica (por exemplo, $\frac{1}{3} = 0,33333...$), na qual alguns números ou uma série de números à direita da vírgula decimal é repetida indefinidamente.

FIGURA 2.1



Uma vez usada a noção de números racionais, surge naturalmente o conceito de *números irracionais* – números que *não podem* ser expressos como razões entre um par de inteiros. Um exemplo é o número $\sqrt{2} = 1,4142\dots$, que é um número decimal não periódico não exato. Um outro exemplo é a constante especial $\pi = 3,1415\dots$ (que representa a razão entre a circunferência de qualquer círculo e seu diâmetro) e que, novamente, é uma decimal não periódica não exata, uma característica de todos os números irracionais.

Se representados sobre uma régua, cada número irracional cairia entre dois números racionais, de modo que, exatamente como as frações preenchem as lacunas entre os números inteiros representados sobre uma régua, os números irracionais preenchem as lacunas entre os números racionais. O resultado desse processo de preenchimento de lacunas é um contínuo de números que, todos juntos, são denominados números reais. Esse contínuo constitui o *conjunto de todos os números reais*, que normalmente é representado pelo símbolo R . Quando o conjunto R é representado sobre uma linha reta (uma régua ampliada), essa linha é denominada *reta real*.

Na Figura 2.1 estão representados todos os conjuntos de números (na ordem discutida) e a relação que guardam uns com os outros. Contudo, lendo a figura de baixo para cima encontraremos, na realidade, um esquema de classificação no qual o conjunto de números reais é dividido em seus conjuntos de números componentes e subcomponentes. Portanto, essa figura é um resumo da estrutura do sistema de números reais.

Para os primeiros 15 capítulos deste livro bastam os números reais, mas eles não são os únicos números utilizados na matemática. Na verdade, a razão para a utilização do termo *real* é que há também números “imaginários”, que têm a ver com as raízes quadradas de números negativos. Esse conceito será discutido mais adiante, no Capítulo 16.

2.3 O conceito de conjunto

Já empregamos a palavra *conjunto* diversas vezes. Considerando que o conceito de conjunto é subjacente a todos os ramos da matemática moderna, é bom que nos familiarizemos pelo menos com seus aspectos mais básicos.

Notação de conjunto

Um *conjunto* é simplesmente uma coleção de objetos distintos. Esses objetos podem ser um grupo de números, pessoas, itens alimentícios ou qualquer outra coisa, distintos. Assim, todos os alunos que freqüentam um determinado curso de economia podem ser considerados um conjunto, exatamente como os três inteiros 2, 3 e 4 podem formar um conjunto. Os objetos em um conjunto são denominados os *elementos* do conjunto.

Há dois modos alternativos de representar um conjunto por escrito: por *enumeração* e por *descrição*. Se S representa o conjunto de números 2, 3 e 4, podemos escrever, por enumeração dos elementos,

$$S = \{2, 3, 4\}$$

Mas, se denominarmos I o conjunto de *todos* os inteiros positivos, a enumeração se torna difícil e, então, poderemos simplesmente descrever os elementos e escrever

$$I = \{x \mid x \text{ um inteiro positivo}\},$$

que se lê da seguinte maneira: “ I é o conjunto de todos os (números) x , tal que x é um número inteiro positivo.” Note que, nos dois casos, utiliza-se um par de chaves para delimitar o conjunto. Na abordagem descritiva sempre se insere uma barra vertical (ou ponto-e-vírgula) para separar o símbolo de designação geral dos elementos da descrição dos elementos. Como outro exemplo, o conjunto de todos os números reais maiores que 2, mas menores que 5 (denominado J) pode ser expresso simbolicamente como

$$J = \{x \mid 2 < x < 5\}$$

Aqui, até mesmo a declaração descritiva é expressada simbolicamente.

Um conjunto com um número finito de elementos, exemplificado pelo conjunto S apresentado anteriormente, é denominado um *conjunto finito*. Por outro lado, os conjuntos I e J , cada um com um número infinito de elementos, são exemplos de *conjuntos infinitos*. Conjuntos finitos são sempre *enumeráveis* (ou *contáveis*), isto é, seus elementos podem ser contados um por um na sequência 1, 2, 3.... Conjuntos infinitos, entretanto, podem ser enumeráveis (conjunto I), ou *não-enumeráveis* (conjunto J). Nesse último caso, não há nenhum modo de associar os elementos do conjunto com os números naturais de contagem 1, 2, 3... e, portanto, o conjunto é não-contável.

O pertencimento a um conjunto é indicado pelo símbolo $\in$ (uma variante da letra grega epsilon, ϵ , para “elemento”), que é lido da seguinte maneira: “é um elemento de”. Assim, para os dois conjuntos S e I definidos anteriormente, podemos escrever

$$2 \in S \quad 3 \in S \quad 8 \in I \quad 9 \in I \quad (\text{etc.})$$

mas, obviamente, $8 \notin S$ (lê-se “8 não é um elemento do conjunto S ”). Se usarmos o símbolo R para denotar o conjunto de todos os números reais, então a afirmação “ x é algum número real” pode ser simplesmente expressa por

$$x \in R$$

Relações entre conjuntos

Quando dois conjuntos são comparados um com o outro, podem ser observados diversos tipos de relações. Se, por acaso, dois conjuntos S_1 e S_2 contiverem elementos idênticos,

$$S_1 = \{2, 7, a, f\} \quad \text{e} \quad S_2 = \{2, a, 7, f\}$$

diz-se que S_1 e S_2 são *iguais* ($S_1 = S_2$). Note que a ordem em que os elementos aparecem em um conjunto não é importante. No entanto, sempre que houver um elemento diferente – mesmo que seja apenas um – em quaisquer dois conjuntos, esses conjuntos não são iguais.

Um outro tipo de relação entre conjuntos é que um conjunto pode ser um *subconjunto* de outro. Se tivermos dois conjuntos

$$S = \{1, 3, 5, 7, 9\} \quad \text{e} \quad T = \{3, 7\}$$

então T é um subconjunto de S porque todo elemento de T também é um elemento de S . Um enunciado mais formal seria: T é um subconjunto de S se, e somente se, $x \in T$ implicar $x \in S$. Utilizando os símbolos de inclusão no conjunto $\subset$ (está contido em) e $\supset$ (inclui), então podemos escrever

$$T \subset S \quad \text{ou} \quad S \supset T$$

É possível que dois dados conjuntos sejam subconjuntos um do outro. Todavia, quando isso ocorre, podemos ter certeza de que esses dois conjuntos são iguais. Enunciando formalmente: podemos ter $S_1 \subset S_2$ e $S_2 \subset S_1$ se, e somente se, $S_1 = S_2$.

Note que, enquanto o símbolo ε está relacionado com um *elemento* individual de um *conjunto*, o símbolo $\subset$ representa um *subconjunto* de um *conjunto*. Aplicando essa idéia podemos declarar, com base na Figura 2.1, que o conjunto de todos os inteiros é um subconjunto do conjunto de todos os números racionais. De modo semelhante, o conjunto de todos os números racionais é um subconjunto do conjunto de todos os números reais.

Quantos subconjuntos podem ser formados com os cinco elementos do conjunto $S = \{1, 3, 5, 7, 9\}$? Antes de mais nada, cada elemento individual de S pode ser contado como um subconjunto distinto de S , tal como $\{1\}$ e $\{3\}$. Mas quaisquer pares, trios ou quadras desses elementos, tais como $\{1, 3\}$, $\{1, 5\}$ e $\{3, 7, 9\}$ também podem ser contados desse mesmo modo. Qualquer subconjunto que *não* contenha *todos* os elementos de S é denominado um *subconjunto próprio* de S . Mas o conjunto S em si (com todos os seus cinco elementos) também pode ser considerado como um de seus próprios subconjuntos – todo elemento de S é um elemento de S e, assim, o próprio conjunto S cumpre a definição de um subconjunto. Evidentemente, este é um caso limite, no qual obtemos o maior subconjunto possível de S , a saber, o próprio conjunto S .

No outro extremo, o menor subconjunto possível de S é um conjunto que não contém absolutamente nenhum elemento. Esse conjunto é denominado *conjunto nulo* ou *conjunto vazio*, representado pelo símbolo $\emptyset$ ou $\{\}$. A razão para considerar o conjunto vazio como um subconjunto de S é bem interessante: se o conjunto nulo não for um subconjunto de S ($\emptyset \not\subset S$), então $\emptyset$ deve conter, no mínimo, um elemento x tal que $x \notin S$. Mas visto que, por definição, o conjunto nulo não tem nenhum elemento sequer, não podemos dizer que $\emptyset \not\subset S$; conseqüentemente, o conjunto nulo é um subconjunto de S .

É extremamente importante distinguir claramente o símbolo $\emptyset$ ou $\{\}$ da notação $\{0\}$; o primeiro não contém nenhum elemento, mas o último contém um elemento: zero. O conjunto nulo é único; há somente um conjunto nulo em todo o mundo e esse conjunto é considerado um subconjunto de *qualquer* conjunto que possa ser concebido.

Contando todos os subconjuntos de S , incluindo os dois casos limites, S e $\emptyset$, encontramos um total de $2^5 = 32$ subconjuntos. Em geral, se um conjunto tiver n elementos, podem ser formados 2^n subconjuntos com esses elementos.[†]

Um terceiro tipo possível de relação entre conjuntos é que dois conjuntos podem não ter absolutamente nenhum elemento em comum. Nesse caso, diz-se que os dois conjuntos são *disjuntos*. Por exemplo, o conjunto de todos os inteiros positivos e o conjunto de todos os inteiros negativos são mutuamente exclusivos; assim, são conjuntos disjuntos.

Um quarto tipo de relação ocorre quando dois conjuntos têm alguns elementos em comum, mas alguns elementos são peculiares a cada um. Quando isso ocorre, os dois conjuntos não são nem iguais nem disjuntos; e, também, nenhum conjunto é um subconjunto do outro.

Operações com conjuntos

Quando somamos, subtraímos, multiplicamos, dividimos ou calculamos a raiz quadrada de alguns números, estamos executando operações matemáticas. Embora conjuntos sejam diferentes de números, também podemos realizar certas operações matemáticas com eles. Três operações principais a serem discutidas aqui são a união, a interseção e o complemento de conjuntos.

Fazer a *união* de dois conjuntos A e B significa formar um novo conjunto contendo aqueles elementos (e somente aqueles elementos) que pertencem a A , a B ou a ambos, A e B . O conjunto de união é simbolizado por $A \cup B$ (lê-se “ A união B ”).

[†] Dado um conjunto com n elementos ($a, b, c, \dots, n$), podemos classificar seus subconjuntos primeiramente em duas categorias: uma contendo o elemento a , e a outra não. Cada uma dessas duas categorias pode, ainda, ser classificada em duas subcategorias: uma contendo o elemento b , e a outra não. Note que, considerando o segundo elemento b , dobramos o número de categorias na classificação de 2 para 4 ($= 2^2$). Pelo mesmo critério, considerar o elemento c aumentará o número total de categorias para 8 ($= 2^3$). Quando todos os n elementos são considerados, o número total de categorias se tornará o número total de subconjuntos e esse número é 2^n .

EXEMPLO 1 Se $A = \{3, 5, 7\}$ e $B = \{2, 3, 4, 8\}$, então

$$A \cup B = \{2, 3, 4, 5, 7, 8\}$$

A propósito, esse exemplo ilustra o caso em que dois conjuntos A e B não são nem iguais nem disjuntos e nenhum é um subconjunto do outro.

EXEMPLO 2 Referindo-nos novamente à Figura 2.1, vemos que a união do conjunto de todos os inteiros e do conjunto de todas as frações é o conjunto de todos os números racionais. Da mesma forma, a união do conjunto dos números racionais com o conjunto dos números irracionais resulta no conjunto de todos os números reais.

A *interseção* de dois conjuntos A e B , por outro lado, é um novo conjunto que contém aqueles elementos (e somente aqueles elementos) pertencentes a *ambos* A e B . O conjunto de interseção é simbolizado por $A \cap B$ (lê-se “ A interseção B ”).

EXEMPLO 3 Considerando os conjuntos A e B do Exemplo 1, podemos escrever

$$A \cap B = \{3\}$$

EXEMPLO 4 Se $A = \{-3, 6, 10\}$ e $B = \{9, 2, 7, 4\}$, então $A \cap B = \emptyset$. O conjunto A e o conjunto B são disjuntos; por conseguinte, sua interseção é o conjunto vazio – nenhum elemento é comum a A e a B .

É óbvio que o conceito de interseção é mais restritivo que o de união. No primeiro, somente os elementos *comuns* a A e B são aceitáveis, ao passo que, no último, pertencer *ou a A ou a B* é suficiente para estabelecer que um elemento pertence ao conjunto união. Portanto, os símbolos operadores $\cap$ e $\cup$ – que, incidentalmente, têm o mesmo tipo de status geral dos símbolos $\sqrt{\quad}$, $+$, $\div$ etc. – têm as conotações “e” e “ou”, respectivamente. Esse ponto pode ser melhor avaliado comparando-se as seguintes definições formais de interseção e união:

Interseção: $A \cap B = \{x \mid x \in A \text{ e } x \in B\}$

União: $A \cup B = \{x \mid x \in A \text{ ou } x \in B\}$

E o *complemento* de um conjunto? Para explicar o que é complemento, em primeiro lugar vamos introduzir o conceito do *conjunto universal*. Em um contexto específico de discussão, se os únicos números usados forem o conjunto dos primeiros sete inteiros positivos, podemos nos referir a ele como o conjunto universal U . Então, tendo um dado conjunto, digamos, $A = \{3, 6, 7\}$, podemos definir um outro conjunto $\tilde{A}$ (lê-se “o complemento de A ”) como o conjunto que contém todos os números do conjunto universal U que não estão no conjunto A . Isto é,

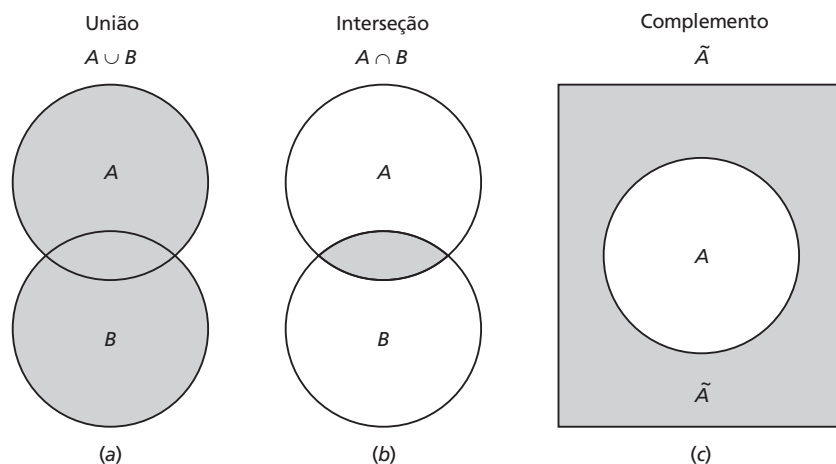


FIGURA 2.2

$$\tilde{A} = \{x \mid x \in U \text{ e } x \notin A\} = \{1, 2, 4, 5\}$$

Note que, enquanto o símbolo $\cup$ tem a conotação “ou” e o símbolo $\cap$ significa “e”, o símbolo de complemento $\sim$ implica a conotação de “não”.

EXEMPLO 5 Se $U = \{5, 6, 7, 8, 9\}$ e $A = \{5, 6\}$, então $\tilde{A} = \{7, 8, 9\}$.

EXEMPLO 6 Qual é o complemento de U ? Visto que cada objeto (número) em consideração está incluído no conjunto universal, o complemento de U deve ser um conjunto vazio. Assim, $\tilde{U} = \emptyset$.

Os três tipos de operações com conjuntos podem ser vistos nos três diagramas da Figura 2.2, conhecidos como *diagramas de Venn*. No diagrama *a*, os pontos no círculo superior formam um conjunto A , e os pontos no círculo inferior formam um conjunto B . Então, a união de A e B consiste na área sombreada que cobre ambos os círculos. No diagrama *b*, são mostrados os mesmos dois conjuntos (círculos). Visto que a interseção desses dois conjuntos deve compreender somente os pontos comuns a ambos os conjuntos, somente a porção sobreposta (sombreada) dos dois círculos satisfaz a definição. No diagrama *c*, sejam os pontos no retângulo o conjunto universal e A o conjunto de pontos no círculo; então, o conjunto complemento $\tilde{A}$ será a área (sombreada) externa ao círculo.

Leis das operações com conjuntos

Podemos notar, na Figura 2.2, que a área sombreada no diagrama *a* representa não somente $A \cup B$, mas também $B \cup A$. Analogamente, no diagrama *b*, a pequena área sombreada é a representação visual não somente de $A \cap B$, mas também de $B \cap A$. Quando formalizado, esse resultado é conhecido como a *lei comutativa* (de uniões e interseções):

$$A \cup B = B \cup A \quad A \cap B = B \cap A$$

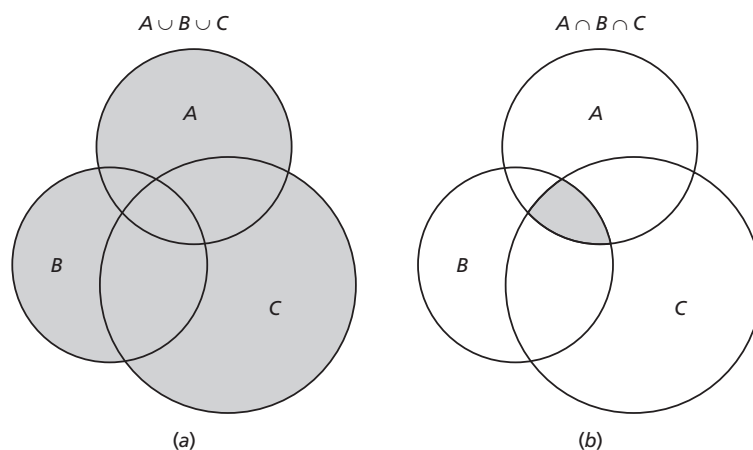
Essas relações são muito semelhantes às leis algébricas $a + b = b + a$ e $a \times b = b \times a$.

Para tomar a união de três conjuntos A , B e C , tomamos primeiramente a união de quaisquer dois conjuntos e, então, a “união” do conjunto resultante com o terceiro; um procedimento semelhante pode ser aplicado à operação de interseção. Os resultados de tais operações estão ilustrados na Figura 2.3. É interessante que a ordem em que esses conjuntos são selecionados para a operação não é importante. Esse fato dá origem à *lei associativa* (de uniões e interseções):

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

FIGURA 2.3



Essas equações lembram muito as leis algébricas $a + (b + c) = (a + b) + c$ e $a \times (b \times c) = (a \times b) \times c$. Há também uma lei de operações que se aplica quando uniões e interseções são usadas combinadamente. É a *lei distributiva* (de uniões e interseções):

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Isso nos lembra a lei algébrica $a \times (b + c) = (a \times b) + (a \times c)$.

EXEMPLO 7 Verifique a lei distributiva dados $A = \{4, 5\}$, $B = \{3, 6, 7\}$ e $C = \{2, 3\}$. Para verificar a primeira parte da lei, em primeiro lugar achamos as expressões à esquerda e à direita separadamente:

Esquerda: $A \cup (B \cap C) = \{4, 5\} \cup \{3\} = \{3, 4, 5\}$

Direita: $(A \cup B) \cap (A \cup C) = \{3, 4, 5, 6, 7\} \cap \{2, 3, 4, 5\} = \{3, 4, 5\}$

Visto que os dois lados dão o mesmo resultado, a lei é verificada. Repetindo o procedimento para a segunda parte da lei, temos

Esquerda: $A \cap (B \cup C) = \{4, 5\} \cap \{2, 3, 6, 7\} = \emptyset$

Direita: $(A \cap B) \cup (A \cap C) = \emptyset \cup \emptyset = \emptyset$

Assim, a lei é novamente verificada.

Verificar uma lei significa aplicá-la a um exemplo específico e verificar se ela realmente funciona. Se a lei for válida, qualquer exemplo específico deverá, de fato, funcionar. Isso implica que, se a lei não for verificada em apenas um único exemplo, ela é inválida. Por outro lado, a verificação bem-sucedida com exemplos específicos (não importa quantos) por si só não prova a lei. Para *provar* uma lei é necessário demonstrar que ela é válida para todos os casos possíveis. O procedimento envolvido nessa demonstração será ilustrado mais adiante (veja, por exemplo, a Seção 2.5).

EXERCÍCIO 2.3

- Escreva o seguinte em notação de conjunto:
 - O conjunto de todos os números reais maiores que 34.
 - O conjunto de todos os números reais maiores que 8, mas menores que 65.
- Dados os conjuntos $S_1 = \{2, 4, 6\}$, $S_2 = \{7, 2, 6\}$, $S_3 = \{4, 2, 6\}$ e $S_4 = \{2, 4\}$, quais das seguintes afirmações são verdadeiras?

(a) $S_1 = S_3$	(d) $3 \notin S_2$	(g) $S_1 \supset S_4$
(b) $S_1 = R$ (conjunto dos números reais)	(e) $4 \notin S_3$	(h) $\emptyset \subset S_2$
(c) $8 \in S_2$	(f) $S_4 \subset R$	(i) $S_3 \supset \{1, 2\}$
- Com referência aos quatro conjuntos dados no Problema 2, encontre:

(a) $S_1 \cup S_2$	(c) $S_2 \cap S_3$	(e) $S_4 \cap S_2 \cap S_1$
(b) $S_1 \cup S_3$	(d) $S_2 \cap S_4$	(f) $S_3 \cap S_1 \cup S_4$
- Quais das seguintes afirmações são válidas?

(a) $A \cup A = A$	(d) $A \cup U = U$	(g) O complemento de $\bar{A}$ é A .
(b) $A \cap A = A$	(e) $A \cap \emptyset = \emptyset$	
(c) $A \cup \emptyset = A$	(f) $A \cap U = A$	
- Dados $A = \{4, 5, 6\}$, $B = \{3, 4, 6, 7\}$ e $C = \{2, 3, 6\}$, verifique a lei distributiva.
- Verifique a lei distributiva por meio de diagramas de Venn, com diferentes ordens sucessivas de sombreamento.
- Enumere todos os subconjuntos do conjunto $\{5, 6, 7\}$.

8. Enumere todos os subconjuntos do conjunto $S = \{a, b, c, d\}$. Quantos subconjuntos há no total?
9. O Exemplo 6 mostra que $\emptyset$ é o complemento de U . Mas, visto que o conjunto nulo é um subconjunto de *qualquer* conjunto, $\emptyset$ deve ser um subconjunto de U . Considerando que o termo “complemento de U ” implica a noção de *não estar em U* , ao passo que o termo “subconjunto de U ” implica a noção de *estar em U* , parece paradoxal que $\emptyset$ seja ambos. Como você resolveria esse paradoxo?

2.4 Relações e funções

Nossa discussão sobre conjuntos foi sugerida pela utilização do termo em conexão com os vários tipos de números de nosso sistema numérico. Contudo, conjuntos também podem se referir a objetos que não sejam números. Em particular, podemos falar de conjuntos de “pares ordenados” – que serão definidos logo a seguir –, que nos levarão aos importantes conceitos de relações e funções.

Pares ordenados

Quando escrevemos um conjunto $\{a, b\}$, não nos importamos com a ordem na qual os elementos a e b aparecem porque, por definição, $\{a, b\} = \{b, a\}$. O par de elementos a e b , nesse caso, é um *par não ordenado*. Quando a ordenação de a e b tem um significado, entretanto, podemos escrever dois *pares ordenados* diferentes denotados por (a, b) e (b, a) , que têm a propriedade de $(a, b) \neq (b, a)$, a não ser que $a = b$. Conceitos semelhantes se aplicam a um conjunto com mais de dois elementos, caso em que podemos distinguir entre triplas, quádruplas, quádruplas ordenadas e não ordenadas e assim por diante. Pares, triplas ordenadas etc. podem ser coletivamente denominadas *conjuntos ordenados* e são limitadas por parênteses, em vez de chaves.

EXEMPLO 1 Para mostrar a idade e o peso de cada aluno em uma sala de aula, podemos formar pares ordenados (a, w) , nos quais o primeiro elemento indica a idade (em anos) e o segundo elemento indica o peso (em libras [em quilogramas]). Então, $(19, 127 [58])$ e $(127 [58], 19)$ obviamente significariam coisas diferentes. Além disso, o último par ordenado dificilmente representaria qualquer estudante em qualquer lugar.

EXEMPLO 2 Quando nos referimos ao conjunto de todos os participantes de uma modalidade dos Jogos Olímpicos, a ordem em que são relacionados não tem nenhuma importância e temos um conjunto não ordenado. Mas o conjunto {medalhista de ouro, medalhista de prata, medalhista de bronze} é uma tripla ordenada.

Pares ordenados, como quaisquer outros objetos, podem ser elementos de um conjunto. Considere o plano retangular de coordenadas (plano cartesiano) na Figura 2.4, no qual um eixo x e um eixo y se cruzam em ângulo reto, dividindo o plano em quatro quadrantes. Esse plano xy é um conjunto infinito de pontos, cada um representando um par ordenado cujo primeiro elemento é um valor x e cujo segundo elemento é um valor y . Claramente, o ponto identificado por $(4, 2)$ é diferente do ponto identificado por $(2, 4)$; portanto, nesse caso, a ordenação é importante.

Como já entendemos essa apresentação visual, estamos prontos para considerar o processo de geração de pares ordenados. Suponha que, a partir de dois conjuntos dados, $x = \{1, 2\}$ e $y = \{3, 4\}$, desejamos formar todos os pares ordenados possíveis tirando o primeiro elemento do conjunto x e o segundo elemento do conjunto y . O resultado será, é claro, o conjunto de pares ordenados $(1, 3)$, $(1, 4)$, $(2, 3)$ e $(2, 4)$. Esse conjunto é denominado o *produto cartesiano* (do nome Descartes), ou *produto direto*, dos conjuntos x e y e é denotado por $x \times y$ (que se lê “ x cruza y ”). É importante lembrar que, conquanto x e y sejam conjuntos de números, o produto cartesiano resulta em um conjunto de pares ordenados. Podemos expressar esse produto cartesiano por enumeração ou por descrição, como:

$$x \times y = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$$

ou
$$x \times y = \{(a, b) \mid a \in x \text{ e } b \in y\}$$

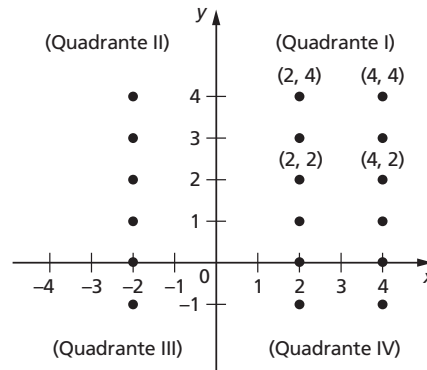


FIGURA 2.4

Aliás, a última expressão pode ser considerada a definição geral de produto cartesiano para quaisquer dados conjuntos x e y .

Para ampliar nosso horizonte, vamos admitir agora que x e y incluem todos os números reais. Então, o produto cartesiano resultante

$$x \times y = \{(a, b) \mid a \in R \text{ e } b \in R\} \quad (2.3)$$

representaria o conjunto de todos os pares ordenados cujos elementos têm valores reais. Além disso, cada par ordenado corresponde a um *único* ponto no plano de coordenadas cartesianas da Figura 2.4 e, inversamente, cada ponto no plano de coordenadas corresponde a um par ordenado *único* no conjunto $x \times y$. Em vista dessa dupla unicidade, diz-se que existe uma *correspondência um a um* entre o conjunto de pares ordenados no produto cartesiano (2.3) e o conjunto de pontos no plano retangular de coordenadas. Agora é fácil de perceber o princípio racional da notação $x \times y$; podemos associá-lo com a interseção do eixo x com o eixo y na Figura 2.4. Um modo mais simples de expressar o conjunto $x \times y$ em (2.3) é escrevê-lo diretamente como $R \times R$, também comumente denotado por R^2 .

Ampliando essa idéia, também podemos definir o produto cartesiano dos três conjuntos x , y e z como segue:

$$x \times y \times z = \{(a, b, c) \mid a \in x, b \in y, c \in z\},$$

que é um conjunto de triplas ordenadas. Além disso, se cada um dos conjuntos x , y e z consistir em todos os números reais, o produto cartesiano corresponderá ao conjunto de todos os pontos em um espaço tridimensional, o que pode ser denotado por $R \times R \times R$ ou, mais simplesmente, R^3 . Na presente discussão, consideramos que todas as variáveis têm valor real; assim, a estrutura será, em termos gerais, R^2 ou R^3 ... ou R^n .

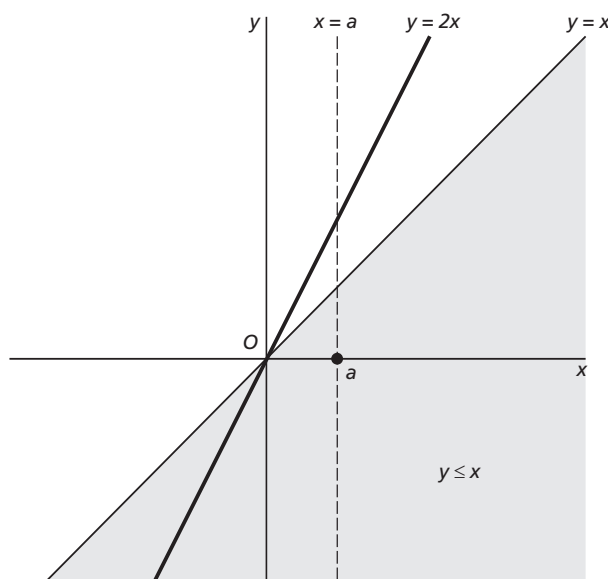
Relações e funções

Uma vez que cada par ordenado associa um valor y com um valor x , qualquer coleção de pares ordenados – qualquer subconjunto do produto cartesiano (2.3) – constituirá uma *relação* entre y e x . Dado um valor x , um ou mais valores y serão especificados por aquela relação. Por conveniência, daqui em diante escreveremos os elementos de $x \times y$ geralmente como (x, y) – em vez de (a, b) , como foi feito em (2.3) – onde ambos, x e y , eram variáveis.

EXEMPLO 3

O conjunto $\{(x, y) \mid y = 2x\}$ é um conjunto de pares ordenados que inclui, por exemplo, $(1, 2)$, $(0, 0)$ e $(-1, -2)$. Ele constitui uma relação, e sua contraparte gráfica é o conjunto de pontos sobre a reta $y = 2x$, mostrado na Figura 2.5.

FIGURA 2.5

**EXEMPLO 4**

O conjunto $\{(x, y) \mid y \leq x\}$, que consiste em pares ordenados como $(1, 0)$, $(1, 1)$ e $(1, -4)$, constitui uma outra relação. Na Figura 2.5, esse conjunto corresponde ao conjunto de todos os pontos na área sombreada que satisfazem a desigualdade $y \leq x$.

Observe que, quando o valor de x é dado, nem sempre é possível determinar um valor *único* de y a partir de uma relação. No Exemplo 4, os três exemplares de pares ordenados mostram que, se $x = 1$, y pode assumir vários valores, tais como 0, 1 ou -4 e, ainda assim, satisfazer, em cada caso, a relação enunciada. Em termos gráficos, dois ou mais pontos de uma relação podem cair sobre uma única reta vertical no plano xy . Isso é exemplificado na Figura 2.5, onde muitos pontos na área sombreada (que representa a relação $y \leq x$) caem na reta vertical tracejada denominada $x = a$.

Todavia, em casos especiais, uma relação pode ser tal que, para cada valor de x , existe somente *um* valor correspondente de y . A relação no Exemplo 3 é um deles. Nesse caso, diz-se que y é uma *função* de x , o que é denotado por $y = f(x)$, que se lê “ y é igual a f de x ”. [Nota: $f(x)$ *não* significa f vezes x .] Uma função é, portanto, um conjunto de pares ordenados que tem a propriedade de qualquer valor de x determinar *unicamente* um valor de y .^{*} Deve ficar claro que uma função tem de ser uma relação, mas uma relação pode não ser uma função.

Embora a definição de uma função estipule um único y para cada x , o inverso não é obrigatório. Em outras palavras, a um mesmo valor y pode ser associado legitimamente mais de um valor x . Essa possibilidade é ilustrada na Figura 2.6, onde os valores x_1 e x_2 no conjunto x são ambos associados com o mesmo valor (y_0) no conjunto y pela função $y = f(x)$.

Uma função também é denominada um *mapeamento* ou *transformação*; ambas as palavras têm como conotação a ação de associar uma coisa com outra. Assim, na declaração $y = f(x)$, a notação funcional f pode ser interpretada como significando uma regra pela qual o conjunto x é “mapeado” (“transformado”) para o (no) conjunto y . Portanto, podemos escrever

$$f: x \rightarrow y$$

onde a seta indica mapeamento e a letra f especifica simbolicamente uma regra de mapeamento. Uma vez que f representa uma regra *particular* de mapeamento, deve ser empregada uma notação funcional diferente para denotar uma outra função que acaso apareça no mesmo modelo. Os símbolos costumeiros (além de f) usados para esse propósito são g , F , G , as letras gregas minúsculas ϕ (fi) e ψ (psi) e suas maiúsculas correspondentes Φ e Ψ . Por exemplo, duas variáveis y e z

^{*} Essa definição de *função* corresponde ao que seria denominado uma *função unívoca* na terminologia mais antiga. O que antes era denominado formalmente uma *função de vários valores* agora é denominado uma *relação* ou correspondência.

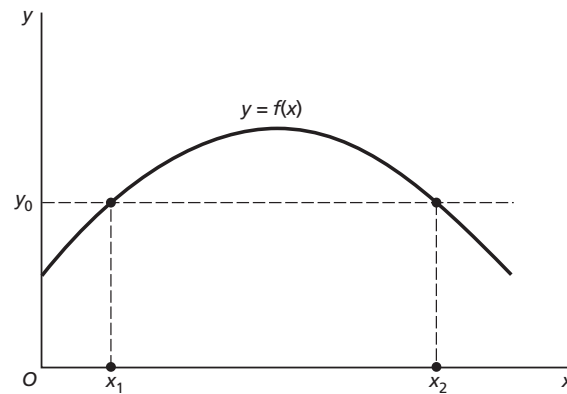


FIGURA 2.6

podem ser, ambas, funções de x , mas, se uma função for escrita como $y = f(x)$, a outra deve ser escrita $z = g(x)$ ou $z = \phi(x)$. Entretanto, também é permitido escrever $y = y(x)$ e $z = z(x)$, dispensando, assim, totalmente, os símbolos f e g .

Na função $y = f(x)$, x é denominado o *argumento* da função e y é denominado o *valor* da função. Vamos nos referir alternativamente a x como a *variável independente* e a y como a *variável dependente*. O conjunto de todos os valores permissíveis que x pode assumir em um dado contexto é conhecido como o *domínio* da função, que pode ser um subconjunto do conjunto de todos os números reais. O valor y para o qual um valor x é mapeado é denominado a *imagem* daquele valor x . O conjunto de todas as imagens é denominado a *imagem* da função, que é o conjunto de todos os valores que a variável y pode assumir. Assim, o domínio pertence à variável independente x , e a imagem tem a ver com a variável independente y .

Como ilustrado na Figura 2.7a, podemos considerar a função f como uma regra para mapear cada ponto sobre algum segmento de reta (o domínio) para algum ponto sobre outro segmento de reta (a imagem). Representando o domínio no eixo x e a imagem no eixo y , como na Figura 2.7b, contudo, obtemos imediatamente o gráfico bidimensional familiar no qual a associação entre valores x e valores y é especificada por um conjunto de pares ordenados como (x_1, y_1) e (x_2, y_2) .

Em modelos econômicos, equações comportamentais usualmente entram como funções. Visto que, por sua própria natureza, quase todas as variáveis em modelos econômicos estão restritas a números reais não negativos,[†] seus domínios também são restritos. É por isso que, em economia, a maioria das representações geométricas é apresentada somente no primeiro quadrante. Em geral, não nos preocuparemos em especificar o domínio de cada função em cada modelo econômico. Quando não for dada nenhuma especificação, deve-se entender que o domínio (e a imagem) incluirá somente números para os quais uma função faz sentido em economia.

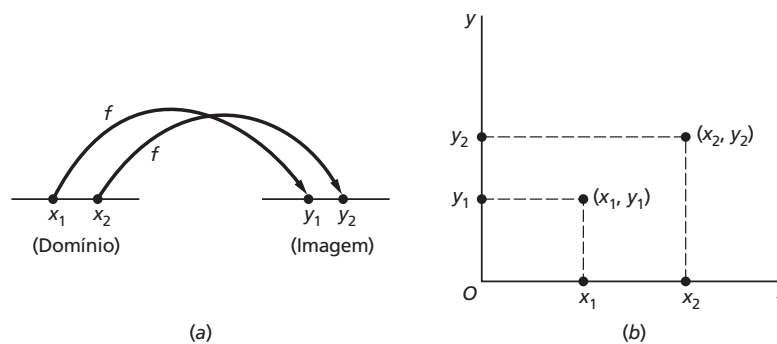


FIGURA 2.7

[†] Dizemos "não negativos" em vez de "positivos" quando são permitidos valores zero.

EXEMPLO 5 O custo total por dia, C , de uma empresa, é uma função de sua produção diária Q : $C = 150 + 7Q$. A empresa tem uma capacidade limite de produção de 100 unidades por dia. Qual é o domínio e a imagem da função custo? Considerando que Q pode variar somente entre 0 e 100, o domínio é o conjunto de valores $0 \leq Q \leq 100$; ou, mais formalmente,

$$\text{Domínio} = \{Q \mid 0 \leq Q \leq 100\}$$

Quanto à imagem, visto que o gráfico da função é uma linha reta cujo valor mínimo de C é 150 (quando $Q = 0$) e cujo valor máximo de C é 850 (quando $Q = 100$), temos

$$\text{Imagem} = \{C \mid 150 \leq C \leq 850\}$$

Porém, cuidado, pois é possível que nem sempre ocorram os valores extremos da faixa quando os valores extremos do domínio forem atingidos.

EXERCÍCIO 2.4

- Dados $S_1 = \{3, 6, 9\}$, $S_2 = \{a, b\}$ e $S_3 = \{m, n\}$, encontre os produtos cartesianos:
 (a) $S_1 \times S_2$ (b) $S_2 \times S_3$ (c) $S_3 \times S_1$
- Com as informações do Problema 1, encontre o produto cartesiano $S_1 \times S_2 \times S_3$.
- Em geral, é verdade que $S_1 \times S_2 = S_2 \times S_1$? Sob que condições esses dois produtos cartesianos serão iguais?
- Algum dos seguintes, desenhado em um plano retangular de coordenadas, representa uma função?
 (a) Um círculo (c) Um retângulo
 (b) Um triângulo (d) Uma linha reta com inclinação descendente
- Se o domínio de uma função $y = 5 + 3x$ for o conjunto $\{x \mid 1 \leq x \leq 9\}$, determine a imagem da função e expresse-a como um conjunto.
- Considerando a função $y = -x^2$, se o domínio for o conjunto de todos os números reais não negativos, qual será sua imagem?
- Na teoria da empresa, economistas consideram que o custo total C é uma função do nível de produção Q : $C = f(Q)$.
 (a) Segundo a definição de uma função, cada número de custo deveria ser associado com um único nível de produção?
 (b) Cada nível de produção determina um único valor do custo?
- Se o nível de produção Q_1 pode ser produzido a um custo C_1 , então também deve ser possível (sendo menos eficiente) produzir Q_1 ao custo $C_1 + \$1$, ou $C_1 + \$2$ e assim por diante. Desse modo, parece que a produção Q não determinaria unicamente o custo total C . Se isso for verdade, escrever $C = f(Q)$ violaria a definição de uma função. Como, a despeito desse raciocínio, você justificaria a utilização da função $C = f(Q)$?

2.5 Tipos de função

A expressão $y = f(x)$ é uma declaração geral no sentido de que um mapeamento é possível, mas isso não quer dizer que a regra de mapeamento propriamente dita fica explicitada. Agora vamos considerar diversos tipos específicos de funções, cada um representando uma regra de mapeamento diferente.

Funções constantes

Uma função cuja faixa consiste em somente um elemento é denominada uma *função constante*. Como exemplo, citamos a função

$$y = f(x) = 7$$

que pode ser expressa alternativamente como $y = 7$ ou $f(x) = 7$, cujo valor permanece o mesmo independentemente do valor de x . No plano cartesiano, tal função apareceria como uma reta horizontal. Em modelos de renda nacional, quando o investimento I é determinado exogenamente, podemos ter uma função de investimento da forma $I = \$100$ milhões ou $I = I_0$, que exemplifica a função constante.

Funções polinomiais

A função constante é, na verdade, um caso “degenerado” das funções conhecidas como *funções polinomiais*. A palavra *polinomial* significa “vários termos” ou “multitermos”, e uma função polinomial de uma única variável x tem a forma geral

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \quad (2.4)$$

na qual cada termo contém um coeficiente, bem como uma potência inteira não negativa da variável x . (Como explicaremos mais adiante nesta seção, em termos gerais, podemos escrever $x^1 = x$ e $x^0 = 1$; assim, os dois primeiros termos podem ser considerados respectivamente como a_0x^0 e a_1x^1). Note que, em vez dos símbolos $a, b, c, \dots$, empregamos os símbolos subscritos $a_0, a_1, \dots, a_n$ para os coeficientes. Isso se deve às duas considerações seguintes: (1) podemos economizar símbolos, já que, desse modo, “esgotamos” somente a utilização da letra a ; e (2) o subscrito ajuda a indicar a localização de um coeficiente específico na equação inteira. Por exemplo, em (2.4), a_2 é o coeficiente de x^2 e assim por diante.

Dependendo do valor do inteiro n (que especifica a potência mais alta de x), temos diversas subclasses de função polinomial:

Caso de $n = 0$:	$y = a_0$	[função <i>constante</i>]
Caso de $n = 1$:	$y = a_0 + a_1x$	[função <i>linear</i>]
Caso de $n = 2$:	$y = a_0 + a_1x + a_2x^2$	[função <i>quadrática</i>]
Caso de $n = 3$:	$y = a_0 + a_1x + a_2x^2 + a_3x^3$	[função <i>cúbica</i>]

e assim por diante. Os indicadores sobrescritos das potências de x são denominados *expoentes*. A potência mais alta envolvida, isto é, o valor de n , é usualmente denominada o *grau* da função polinomial; uma função quadrática, por exemplo, é um polinômio de segundo grau e uma função cúbica é um polinômio de terceiro grau.[†] A ordem em que os diversos termos aparecem à direita do sinal de igual não é importante; eles também poderiam estar organizados em ordem decrescente de potência. Além disso, mesmo que tenhamos escrito o símbolo y à esquerda do sinal de igual, também é aceitável escrever $f(x)$ em seu lugar.

Uma função linear, quando representada graficamente no plano cartesiano, terá a forma de uma linha reta, como mostra a Figura 2.8a. Quando $x = 0$, a função linear resulta em $y = a_0$; assim, o par ordenado $(0, a_0)$ está sobre a reta. Isso nos dá o que denominamos a interseção de y (ou *interseção vertical*) porque é nesse ponto que o eixo vertical intercepta a reta. O outro coeficiente, a_1 , mede a *inclinação* (o grau de inclinação) da nossa reta. Isso significa que o aumento de uma unidade em x resultará em um incremento em y igual à quantidade a_1 . A Figura 2.8a ilustra o caso de $a_1 > 0$, que envolve uma inclinação positiva e, assim, uma reta de inclinação ascendente; se $a_1 < 0$, a reta terá uma inclinação decrescente.

A representação gráfica de uma função quadrática, por outro lado, é uma *parábola* – grosso modo, uma curva que contém uma única concavidade. A ilustração específica na Figura 2.8b implica um a_2 negativo; no caso de $a_2 > 0$, a curva se “abrirá” para o outro lado, apresentando um

[†] Nas diversas equações que acabamos de citar, sempre se supõe que o coeficiente de grau mais alto (a_n) é diferente de zero (não-zero); caso contrário, a função degeneraria para um polinômio de grau mais baixo.

vale em vez de uma elevação. O gráfico de uma função cúbica em geral apresenta duas concavidades, como ilustra a Figura 2.8c. Essas funções serão usadas com bastante frequência nos modelos econômicos que serão discutidos mais adiante.

Funções racionais

Uma função como

$$y = \frac{x - 1}{x^2 + 2x + 4}$$

na qual y é expresso como uma razão de dois polinômios da variável x , é conhecida como uma *função racional*. Segundo essa definição, qualquer função polinomial deve ser, em si, uma função racional, porque sempre pode ser expressa como uma razão com 1 e 1 é uma função constante.

Uma função racional especial que tem aplicações interessantes na economia é a função

$$y = \frac{a}{x} \quad \text{ou} \quad xy = a$$

cujas representação gráfica é uma *hipérbole retangular*, como na Figura 2.8d. Visto que, nesse caso, o produto das duas variáveis é sempre uma constante fixa, essa função pode ser utilizada para representar aquela curva especial de demanda – cujos dois eixos são o preço P e a quantidade Q – para a qual o dispêndio total PQ é constante em todos os níveis de preços. (Essa é a curva de demanda cuja elasticidade é unitária em cada ponto da curva.) Uma outra aplicação é à curva do custo fixo médio (*average fixed cost* – AFC). Representando o AFC em um eixo e a produção Q em outro, a curva AFC deve ser uma *hipérbole retangular* porque $\text{AFC} \times Q$ (= custo fixo total) é uma constante fixa.

A *hipérbole retangular* obtida de $xy = a$ nunca toca os eixos, mesmo que estendida indefinidamente para cima e para a direita. Em vez disso, a curva se aproxima *assintoticamente* dos eixos: à medida que y se torna muito grande, a curva fica cada vez mais próxima do eixo y mas nunca o alcança, o mesmo acontecendo com o eixo x . Os eixos constituem as *assíntotas* dessa função.

Funções não-algébricas

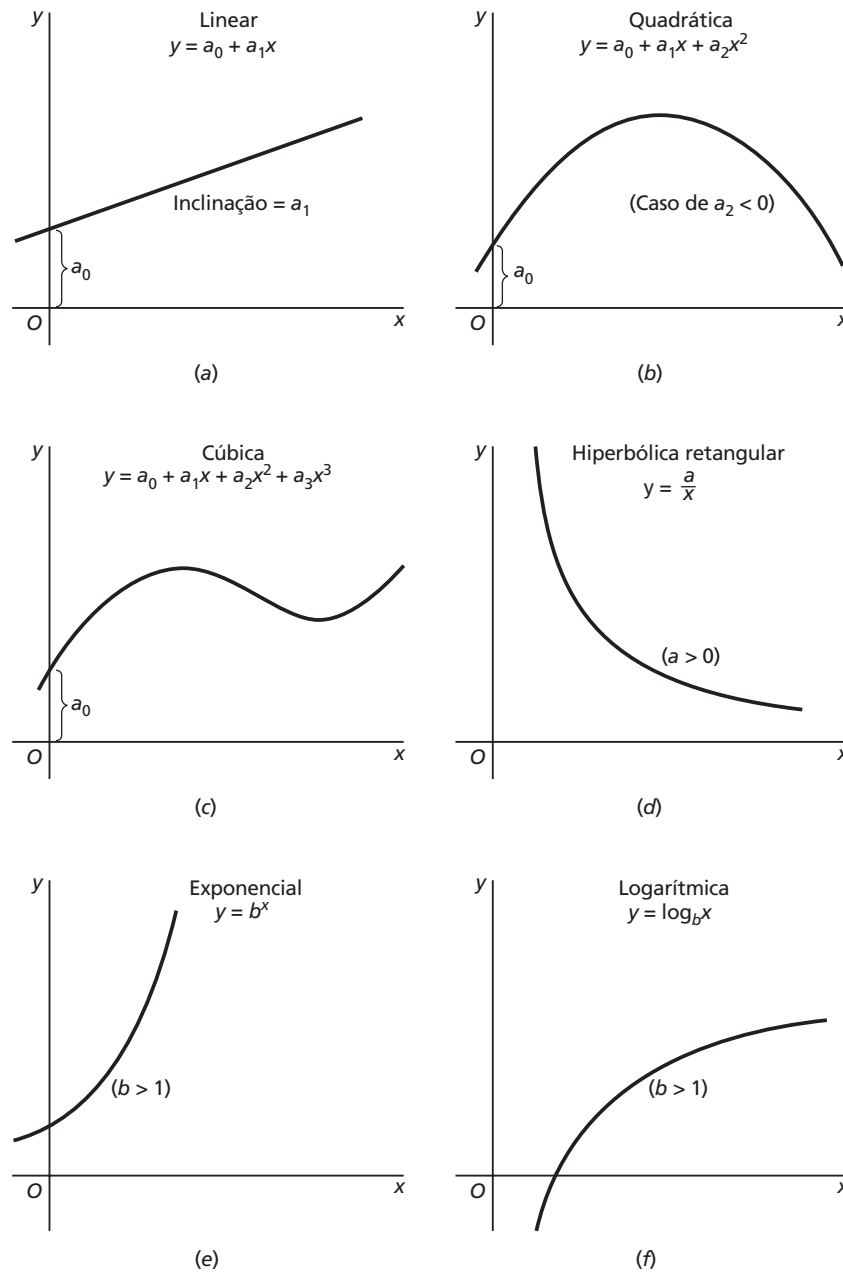
Qualquer função expressa em termos de polinômios e/ou raízes (tal como raiz quadrada) de polinômios é uma *função algébrica*. Segundo essa definição, as funções discutidas até aqui são todas algébricas.

Todavia, *funções exponenciais* como $y = b^x$, nas quais a variável independente aparece no expoente, são *funções não-algébricas*. As *funções logarítmicas*, tais como $y = \log_b x$, estreitamente relacionadas, também são não-algébricas. Esses dois tipos de funções têm um papel especial a desempenhar em certos tipos de aplicações econômicas e, do ponto de vista pedagógico, convém adiar sua discussão para o Capítulo 10. Aqui, apresentaremos simplesmente suas formas gráficas gerais nas Figuras 2.8e e 2.8f. Outros tipos de funções não-algébricas são as *funções trigonométricas* (ou *circulares*), que discutiremos no Capítulo 16, em conexão com análise dinâmica. Devemos acrescentar aqui que funções não-algébricas também são conhecidas pelo nome mais esotérico de *funções transcendentais*.

Uma digressão sobre expoentes

Quando discutimos funções polinomiais, introduzimos o termo *expoentes* como indicadores da potência à qual uma variável (ou número) deve ser elevada. A expressão 6^2 significa que 6 deve ser elevado à segunda potência; isto é, 6 deve ser multiplicado por si mesmo, ou $6^2 = 6 \times 6 = 36$. Em geral, definimos, para um inteiro positivo n ,

FIGURA 2.8



$$x^n \equiv \underbrace{x \times x \times \cdots \times x}_{n \text{ termos}}$$

e, como um caso especial, observamos que $x^1 = x$. Da definição geral, segue que, para inteiros positivos m e n , os expoentes obedecem às seguintes regras:

Regra I

$$x^m \times x^n = x^{m+n} \quad (\text{por exemplo, } x^3 \times x^4 = x^7)$$

DEMONSTRAÇÃO

$$\begin{aligned} x^m \times x^n &= \underbrace{(x \times x \times \cdots \times x)}_{m \text{ termos}} \underbrace{(x \times x \times \cdots \times x)}_{n \text{ termos}} \\ &= \underbrace{x \times x \times \cdots \times x}_{m+n \text{ termos}} = x^{m+n} \end{aligned}$$

Note que, nessa demonstração, não atribuímos nenhum valor específico ao número x ou aos expoentes m e n . Assim, o resultado obtido é verdadeiro *em geral*. É por essa razão que o argumento apresentado constitui uma demonstração, ao contrário de uma mera verificação. O mesmo pode ser dito sobre a Regra II a seguir:

$$\text{Regra II} \quad \frac{x^m}{x^n} = x^{m-n} \quad (x \neq 0) \quad \left(\text{por exemplo, } \frac{x^4}{x^3} = x \right)$$

$$\text{DEMONSTRAÇÃO} \quad \frac{x^m}{x^n} = \frac{\overbrace{x \times x \times \cdots \times x}^{m \text{ termos}}}{\underbrace{x \times x \times \cdots \times x}_n} = \underbrace{x \times x \times \cdots \times x}_{m-n \text{ termos}} = x^{m-n}$$

porque os n termos no denominador cancelam n dos m termos no numerador. Note que o enunciado desta regra exclui o caso de $x = 0$. Isso porque, quando $x = 0$, a expressão x^m/x^n envolveria uma divisão por zero, que é indefinida.

E se $m < n$, digamos, $m = 2$ e $n = 5$? Nesse caso, teremos, segundo a Regra II, $x^{m-n} = x^{-3}$, uma *potência negativa* de x . O que isso significa? Na verdade, a resposta é dada pela própria Regra II: quando $m = 2$ e $n = 5$, temos

$$\frac{x^2}{x^5} = \frac{x \times x}{x \times x \times x \times x \times x} = \frac{1}{x \times x \times x} = \frac{1}{x^3}$$

Assim, $x^{-3} = 1/x^3$ e isso pode ser generalizado em uma outra regra:

$$\text{Regra III} \quad x^{-n} = \frac{1}{x^n} \quad (x \neq 0)$$

Elevar um número (diferente de zero) a uma potência n *negativa* é tomar o inverso de sua *enésima* potência.

Um outro caso especial na aplicação da Regra II é quando $m = n$, o que resulta na expressão $x^{m-n} = x^{m-m} = x^0$. Para interpretar o significado de elevar um número x à potência zero, podemos escrever o termo x^{m-m} conforme a Regra II, cujo resultado é $x^m/x^m = 1$. Assim, podemos concluir que qualquer número (diferente de zero) elevado à potência zero é igual a 1. (A expressão 0^0 é indefinida). Isso pode ser expresso por uma outra regra:

$$\text{Regra IV} \quad x^0 = 1 \quad (x \neq 0)$$

Enquanto tratarmos somente de funções polinomiais, são necessárias apenas potências inteiras (não-negativas). Em funções exponenciais, contudo, o expoente é uma variável que pode assumir também valores não inteiros. Para representar um número como $x^{1/2}$, vamos considerar o fato de termos, pela Regra I,

$$x^{1/2} \times x^{1/2} = x^1 = x$$

Uma vez que $x^{1/2}$ multiplicado por si mesmo é x , $x^{1/2}$ deve ser a raiz quadrada de x . De modo semelhante, pode-se demonstrar que $x^{1/3}$ é a raiz cúbica de x . Por conseguinte, em termos gerais, podemos enunciar a seguinte regra:

$$\text{Regra V} \quad x^{1/n} = \sqrt[n]{x}$$

Duas outras regras obedecidas por expoentes são:

Regra VI

$$(x^m)^n = x^{mn}$$

Regra VII

$$x^m \times y^m = (xy)^m$$

EXERCÍCIO 2.5

- Represente graficamente as funções abaixo:
 (a) $y = 16 + 2x$ (b) $y = 8 - 2x$ (c) $y = 2x + 12$
 (Considere, em cada caso, que o domínio consiste apenas em números reais não-negativos.)
- Qual é a principal diferença entre (a) e (b) no Problema 1? Como essa diferença se reflete nos gráficos? Qual é a principal diferença entre (a) e (c)? Como seus gráficos a refletem?
- Represente graficamente as funções abaixo:
 (a) $y = -x^2 + 5x - 2$ (b) $y = x^2 + 5x - 2$
 tendo como domínio o conjunto de valores $-5 \leq x \leq 5$. Todos sabem que o sinal do coeficiente do termo x^2 determina se o gráfico de uma função quadrática terá uma "colina" ou um "vale". Tomando como base o presente problema, qual sinal estará associado a uma colina? Dê uma explicação intuitiva para isso.
- Represente graficamente a função $y = 36/x$, admitindo que x e y podem assumir somente valores positivos. Em seguida, suponha que ambas as variáveis podem assumir valores negativos também; como o gráfico deve ser modificado para refletir essa mudança na premissa?
- Condense as seguintes expressões:
 (a) $x^4 \times x^{15}$ (b) $x^a \times x^b \times x^c$ (c) $x^3 \times y^3 \times z^3$
- Encontre:
 (a) x^3/x^{-3} (b) $(x^{1/2} \times x^{1/3})/x^{2/3}$
- Mostre que $x^{m/n} = \sqrt[n]{x^m} = (\sqrt[n]{x})^m$. Especifique as regras aplicadas em cada etapa.
- Demonstre a Regra VI e a Regra VII.

2.6 Funções de duas ou mais variáveis independentes

Até aqui, consideramos somente funções de uma única variável independente, $y = f(x)$. Mas o conceito de uma função pode ser prontamente estendido para o caso de duas ou mais variáveis independentes. Dada a função

$$z = g(x, y)$$

um dado par de valores x e y determinará unicamente um valor da variável dependente z . Tal função é exemplificada por

$$z = ax + by \quad \text{ou} \quad z = a_0 + a_1x + a_2x^2 + b_1y + b_2y^2$$

Exatamente como a função $y = f(x)$ mapeia um ponto no domínio em um ponto da imagem, a função g fará precisamente o mesmo. Contudo, neste caso, o domínio não é mais um conjunto de números, mas um conjunto de pares ordenados (x, y) , porque só podemos determinar z quando *ambos*, x e y , forem especificados. Assim, a função g é um mapeamento de um ponto em um espaço bidimensional para um ponto sobre um segmento de reta (isto é, um ponto que está em um espaço unidimensional), tal como do ponto (x_1, y_1) para o ponto z_1 , ou do ponto (x_2, y_2) para z_2 na Figura 2.9a.

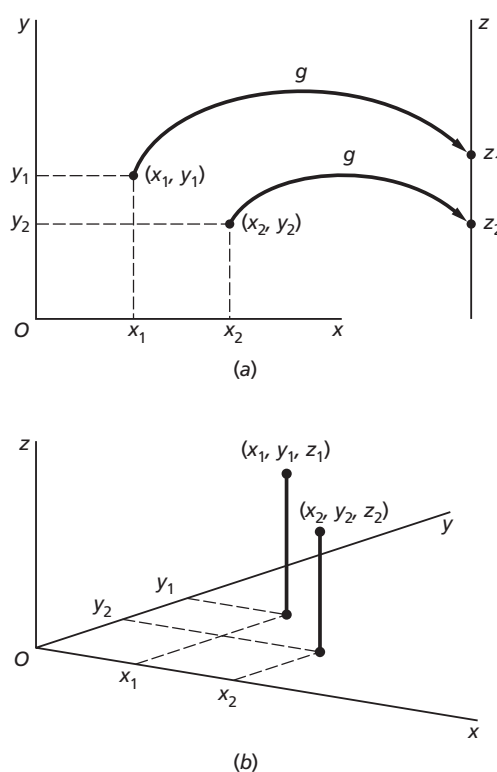
No entanto, se acrescentarmos um eixo vertical z perpendicular ao plano xy , como foi feito no diagrama *b*, o resultado será um espaço tridimensional no qual a função g pode ser representada graficamente como descrito a seguir. O domínio da função será algum subconjunto dos pontos no plano xy , e o valor da função (valor de z) para um dado ponto no domínio – digamos (x_1, y_1) – pode ser indicado pela altura de uma linha vertical desenhada naquele ponto. Assim, a associação entre as três variáveis é resumida pela tripla ordenada (x_1, y_1, z_1) , que é um ponto específico no espaço tridimensional. O locus dessas triplas ordenadas, que tomará a forma de uma *superfície*, constitui, então, o gráfico da função g . Enquanto a função $y = f(x)$ é um conjunto de *pares* ordenados, a função $z = g(x, y)$ será um conjunto de *triplos* ordenados. Haverá muitas ocasiões para usar funções desse tipo em modelos econômicos. Uma aplicação imediata está na área de funções de produção. Suponha que a produção seja determinada por quantidades de capital (K) e mão-de-obra (L) empregadas; então, podemos escrever uma função de produção na forma geral $Q = Q(K, L)$.

A possibilidade de extensão adicional para os casos de três ou mais variáveis independentes agora é evidente por si só. Com a função $y = h(u, v, w)$, por exemplo, podemos mapear um ponto no espaço tridimensional (u, v, w) em um ponto em um espaço unidimensional (y). Uma função como essa poderia ser utilizada para indicar que a utilidade de um consumidor é uma função de seu consumo de três mercadorias diferentes, e o mapeamento em questão é de um espaço tridimensional de mercadoria para um espaço unidimensional de utilidade. Mas, dessa vez, será impossível representar a função graficamente porque seria preciso um diagrama de quatro dimensões para representar as quádruplas ordenadas, mas o mundo em que vivemos é apenas tridimensional. Não obstante, em vista do apelo intuitivo da analogia geométrica, podemos continuar a denominar uma quádrupla ordenada (u, v, w, y) como um “ponto” no espaço de quatro dimensões. O locus desses pontos resultará no “gráfico” (impossível de representar graficamente) da função $y = h(u, v, w)$, que é denominada uma *hipersuperfície*. Esses termos, a saber, ponto e hipersuperfície, também se aplicam ao caso geral do espaço de n dimensões.

Funções de mais de uma variável também podem ser classificadas em vários tipos. Por exemplo, uma função da forma

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

FIGURA 2.9



é uma função *linear* cuja característica é que cada variável é elevada somente à primeira potência. Uma função *quadrática*, por outro lado, envolve primeiras e segundas potências de uma ou mais variáveis independentes, mas a soma dos expoentes das variáveis que aparecem em cada termo não deve ser maior que 2.

Observe que, em vez de denotar as variáveis independentes x, u, v, w etc., passamos para os símbolos $x_1, x_2, \dots, x_n$. Essa última notação, como o sistema de coeficientes subscritos, tem o mérito de economizar o alfabeto, bem como de proporcionar uma percepção mais fácil do número de variáveis envolvidas em uma função.

2.7 Níveis de generalidade

Quando discutimos os vários tipos de funções, apresentamos, sem nenhuma observação explícita, exemplos de funções que pertencem a níveis variados de generalidade. Em certas instâncias, escrevemos funções na forma

$$y = 7 \quad y = 6x + 4 \quad y = x^2 - 3x + 1 \quad (\text{etc.})$$

Essas funções não somente são expressas em termos de coeficientes numéricos, mas também indicam especificamente se cada uma delas é constante, linear ou quadrática. Em termos de gráficos, cada uma dessas funções dará origem a uma curva única, bem definida. Em vista da natureza numérica dessas funções, as soluções do modelo baseado nelas também serão valores numéricos. A desvantagem é que, cada vez que quisermos saber como nossa conclusão analítica mudará quando entrar em jogo um conjunto diferente de coeficientes numéricos, teremos de percorrer todo o processo de raciocínio novamente. Assim, os resultados obtidos de funções específicas não têm muita generalidade.

Passando para um nível mais geral de discussão e análise, há funções na forma

$$y = a \quad y = a + bx \quad y = a + bx + cx^2 \quad (\text{etc.})$$

Uma vez que são utilizados parâmetros, cada função representa não uma única curva, mas toda uma família de curvas. A função $y = a$, por exemplo, abrange não somente os casos específicos $y = 0, y = 1$ e $y = 2$, mas também $y = \frac{1}{3}, y = -5, \dots$, até o infinito. Com funções paramétricas, o resultado de operações matemáticas também será em termos de parâmetros. Esses resultados são mais gerais no sentido de que, atribuindo vários valores aos parâmetros que aparecem na solução do modelo, podemos obter toda uma família de respostas específicas sem ter de repetir o processo de raciocínio novamente.

Para atingir um nível de generalidade mais alto ainda, podemos recorrer ao enunciado geral $y = f(x)$ ou $z = g(x, y)$. Quando expressa dessa forma, a função não fica restrita a ser exclusivamente linear, quadrática, exponencial ou trigonométrica – todas são incorporadas na notação. A aplicabilidade do resultado analítico baseado em tal formulação geral será, portanto, a mais geral possível. Contudo, como veremos mais adiante, para obter resultados economicamente significativos, muitas vezes é necessário impor certas restrições qualitativas às funções gerais embutidas em um modelo, tal como a restrição de que uma função de demanda tenha um gráfico de inclinação negativa ou que a função de consumo tenha um gráfico com uma inclinação positiva menor que 1.

Resumindo o presente capítulo, a estrutura de um modelo matemático econômico agora está clara. Em geral, consistirá em um sistema de equações, que podem ser definicionais, comportamentais ou da natureza de condições de equilíbrio.[†] As equações comportamentais usualmente estão sob a forma de funções, que podem ser lineares ou não lineares, numéricas ou paramétricas, e ter uma variável independente ou muitas. É por meio delas que é dada expressão matemática às premissas analíticas adotadas no modelo.

[†] Desigualdades também podem entrar como um componente importante de um modelo, mas não nos preocuparemos com elas por enquanto.

Ao atacar um problema analítico, portanto, o primeiro passo é selecionar as variáveis adequadas – exógenas, bem como endógenas – para inclusão no modelo. Em seguida, devemos traduzir em equações o conjunto de premissas analíticas referentes aos aspectos ambientais humanos, institucionais, tecnológicos, legais e outros, que afetam o funcionamento das variáveis. Somente então podemos tentar obter um conjunto de conclusões por meio de operações e manipulações matemáticas relevantes e dar-lhes interpretações econômicas apropriadas.

PARTE 2

Análise estática (ou de equilíbrio)



CAPÍTULO 3

Análise de equilíbrio em economia

O procedimento analítico esboçado no Capítulo 2 será aplicado, em primeiro lugar, ao que é conhecido como *análise estática* ou *análise de equilíbrio*. Para tal finalidade, é imperativo, antes de mais nada, entender claramente o que significa *equilíbrio*.

3.1 O significado de equilíbrio

Como qualquer termo utilizado em economia, *equilíbrio* pode ser definido de diversas maneiras. Segundo uma das definições, um equilíbrio é “uma constelação de variáveis inter-relacionadas selecionadas, ajustadas umas às outras de tal maneira que nenhuma tendência inerente à mudança prevaleça no modelo que elas constituem”.[†] Diversas palavras nessa definição merecem especial atenção. Em primeiro lugar, a palavra *selecionada* salienta o fato de existirem variáveis que, por escolha do analista, não foram incluídas no modelo. Conseqüentemente, o equilíbrio em questão pode ter relevância somente no contexto do conjunto particular de variáveis escolhidas e, se o modelo for ampliado para incluir variáveis adicionais, o estado de equilíbrio pertinente ao modelo menor não se aplicará mais.

Em segundo lugar, a palavra *inter-relacionada* indica que, para ocorrer o equilíbrio, todas as variáveis no modelo deverão estar simultaneamente em um estado de repouso. Além disso, o estado de repouso de cada variável deve ser compatível com o estado de repouso de cada uma das outras variáveis; caso contrário, uma ou algumas dessas variáveis estará mudando e, conseqüentemente, por uma reação em cadeia, também estará fazendo com que outras variáveis mudem e, então, pode-se dizer que não existe equilíbrio algum.

Em terceiro lugar, a palavra *inerente* implica que, ao definir um equilíbrio, o estado de repouso envolvido é baseado apenas no equilíbrio das forças internas do modelo, admitindo-se, ao mesmo tempo, que os fatores externos são fixos. Em termos operacionais, isso significa que parâmetros e variáveis exógenas são tratados como constantes. Quando os fatores externos mudarem, haverá um novo equilíbrio definido com base nos valores dos novos parâmetros, porém, ao definir o novo equilíbrio, admite-se novamente que os valores dos novos parâmetros persistem e permanecem imutáveis.

[†] Machlup., Fritz. “Equilibrium and Disequilibrium: Misplaced Concreteness and Disguised Politics”. *Economic Journal*, p. 9, mar. 1958. (Publicado novamente em Machlup., F. *Essays on Economic Semantics*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1963.)

Em essência, um equilíbrio para um modelo especificado é uma situação caracterizada por uma ausência de tendência à mudança. É por essa razão que a análise de equilíbrio (mais especificamente o estudo de como é o estado de equilíbrio) é denominada *estática*.

O fato de um equilíbrio implicar nenhuma tendência à mudança pode nos tentar a concluir que ele constitui necessariamente um estado de coisas desejável ou ideal, tendo como premissa que somente no estado ideal haveria uma ausência de motivação para mudança. Essa conclusão é injustificável. Mesmo que uma determinada posição de equilíbrio possa representar um estado desejável e algo por que se deva lutar – tal como uma situação de maximização do lucro, do ponto de vista da empresa – uma outra posição de equilíbrio talvez seja bastante indesejável e, portanto, algo que deve ser evitado, tal como um nível de equilíbrio de subutilização da renda nacional. A única interpretação justificável é que um equilíbrio é uma situação que, se atingida, tenderia a se perpetuar, salvo quaisquer mudanças nas forças externas.

A variedade desejável de equilíbrio, a qual denominaremos *equilíbrio-alvo*, será discutida mais adiante, na Parte 4, como problemas de otimização. Neste capítulo, a discussão será restrita ao tipo de equilíbrio que não tem *um alvo definido*, que não resulta de nenhuma intenção consciente de atingir um determinado objetivo, mas de um processo impessoal ou suprapessoal de interação e ajuste de forças econômicas. Exemplos desse tipo de equilíbrio são o equilíbrio atingido por um mercado sob determinadas condições de demanda e oferta e o equilíbrio da renda nacional sob determinadas condições de consumo e padrões de investimento.

3.2 Equilíbrio parcial de mercado – um modelo linear

Em um modelo de equilíbrio estático, o problema padrão é achar o conjunto de valores das variáveis endógenas que satisfarão a condição de equilíbrio do modelo. Isso porque, uma vez identificados esses valores, teremos, na verdade, identificado o estado de equilíbrio. Vamos ilustrar com o que denominamos um modelo de equilíbrio parcial de mercado, isto é, um modelo de determinação de preço em um mercado isolado.

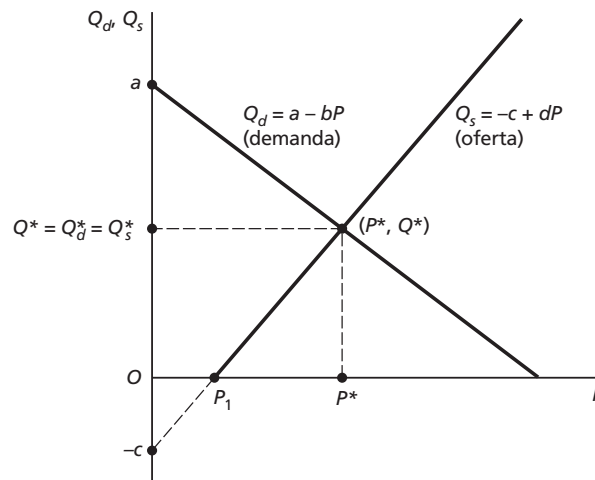
Construção do modelo

Uma vez que somente uma mercadoria está sendo considerada, é necessário incluir somente três variáveis no modelo: a quantidade demandada da mercadoria (Q_d), a quantidade ofertada da mercadoria (Q_s) e seu preço (P). A quantidade é medida, por exemplo, em quilos por semana, e o preço em reais. Escolhidas as variáveis, nossa próxima tarefa é adotar certas premissas em relação ao funcionamento do mercado. Em primeiro lugar, devemos especificar uma condição de equilíbrio – algo indispensável em um modelo de equilíbrio. A premissa padrão é que ocorre equilíbrio no mercado se, e somente se, o excesso de demanda for zero ($Q_d - Q_s = 0$), isto é, se, e somente se, o mercado estiver servido. Porém, isso levanta imediatamente a questão do modo como as próprias Q_d e Q_s são determinadas. Para responder a essa pergunta, admitimos que Q_d é uma função linear decrescente de P (à medida que P aumenta, Q_d diminui). Por outro lado, Q_s é postulada como uma função linear crescente de P (à medida que P aumenta, Q_s também aumenta), com a condição de que nenhuma quantidade é ofertada a menos que o preço exceda um determinado nível positivo. Então, ao todo, o modelo conterá uma condição de equilíbrio e mais duas equações comportamentais que regerão os lados da demanda e da oferta do mercado, respectivamente.

Traduzido para enunciados matemáticos, o modelo pode ser escrito como

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= a - bP & (a, b > 0) \\ Q_s &= -c + dP & (c, d > 0) \end{aligned} \tag{3.1}$$

FIGURA 3.1



Quatro parâmetros, a , b , c e d , aparecem nas duas funções lineares, e todos eles são especificados como positivos. Quando a função de demanda é representada graficamente, como na Figura 3.1, ela intercepta o eixo vertical em a , e sua inclinação é $-b$, negativa, como requerido. A função de oferta também tem o tipo requerido de inclinação, já que d é positivo, mas vemos que sua interseção com o eixo vertical é negativa, em $-c$. Por que quisemos especificar tal interseção vertical negativa? A resposta é que, ao fazer isso, obrigamos a curva de oferta a ter uma interseção horizontal positiva em P_1 , satisfazendo, desse modo, a condição determinada anteriormente de que não ocorrerá oferta a menos que o preço seja positivo e suficientemente alto.

O leitor deve observar que, ao contrário da prática usual, a quantidade, e não o preço, foi colocada no eixo vertical na Figura 3.1. Contudo, isso segue a convenção matemática de colocar a variável *dependente* no eixo vertical. Em um contexto diferente, no qual, do ponto de vista de uma empresa comercial, se considera que a curva de demanda descreve a curva da receita média, $AR \equiv P = f(Q_d)$, invertemos os eixos e colocaremos P no eixo vertical.

Agora que o modelo foi construído, a próxima etapa é resolvê-lo, isto é, obter os valores de solução das três variáveis endógenas, Q_d , Q_s e P . Os valores de solução são aqueles que satisfazem as três equações em (3.1) simultaneamente; isto é, são os valores que, quando substituídos nas três equações, fazem delas um conjunto de enunciados verdadeiros. No contexto de um modelo de equilíbrio, esses valores também podem ser denominados *valores de equilíbrio* das variáveis citadas.

Há muitos autores que não empregam símbolos especiais para denotar os valores de solução das variáveis endógenas. Assim, Q_d é usado para representar a variável quantidade demandada (com toda uma faixa de valores) ou o valor da solução (um valor específico), o mesmo acontecendo com os símbolos Q_s e P . Infelizmente, essa prática pode dar origem a possíveis confusões, especialmente no contexto da análise estática comparativa (ver Seção 7.5). Para evitar tal fonte de confusão, denotaremos o valor da solução de uma variável endógena com um asterisco. Assim, os valores de solução de Q_d , Q_s e P são denotados por Q_d^* , Q_s^* e P^* , *respectivamente*. Porém, visto que $Q_d^* = Q_s^*$, esses símbolos podem até mesmo ser substituídos por um único símbolo Q^* . Por conseguinte, uma solução de equilíbrio do modelo pode ser denotada simplesmente por um par ordenado (P^*, Q^*) . No caso de a solução não ser única, diversos pares ordenados podem satisfazer o sistema de equações simultâneas; então haverá um conjunto de soluções que contém mais de um elemento. Todavia, a situação de equilíbrio múltiplo não pode surgir em um modelo linear tal como o presente.

Solução por eliminação de variáveis

Um modo de achar uma solução para um sistema de equações é pela eliminação sucessiva de variáveis e equações por substituição. Em (3.1), o modelo contém três equações e três variáveis. Contudo, em vista da igualdade entre Q_d e Q_s determinada pela condição de equilíbrio, podemos deixar $Q = Q_d = Q_s$ e reescrever o modelo na seguinte forma equivalente:

$$\begin{aligned} Q &= a - bP \\ Q &= -c + dP \end{aligned} \quad (3.2)$$

o que o reduz a duas equações com duas variáveis. Além disso, substituindo a primeira equação na segunda em (3.2), o modelo pode ser reduzido ainda mais, resultando em uma única equação, com uma única variável:

$$a - bP = -c + dP$$

ou, após subtrair $(a + dP)$ de ambos os lados da equação e multiplicar tudo por -1 ,

$$(b + d)P = a + c \quad (3.3)$$

Esse resultado também pode ser obtido diretamente de (3.1) substituindo a segunda e a terceira equações na primeira.

Visto que $b + d \neq 0$, podemos dividir ambos os lados de (3.3) por $(b + d)$. O resultado é o valor da solução de P :

$$P^* = \frac{a + c}{b + d} \quad (3.4)$$

Note que P^* é – como todos os valores da solução devem ser – expresso integralmente em termos dos parâmetros, que representam dados determinados para o modelo. Assim, P^* é um valor determinado, como deve ser. Note também que P^* é positivo – como um preço deve ser – porque todos os quatro parâmetros são positivos por especificação do modelo.

Para achar a quantidade de equilíbrio Q^* ($= Q_d^* = Q_s^*$) que corresponde ao valor P^* , basta substituir (3.4) em *qualquer* equação de (3.2) e resolver a equação resultante. Substituindo (3.4) na função de demanda, por exemplo, podemos obter

$$Q^* = a - \frac{b(a + c)}{b + d} = \frac{a(b + d) - b(a + c)}{b + d} = \frac{ad - bc}{b + d} \quad (3.5)$$

que, novamente, é uma expressão em termos de parâmetros, apenas. Como o denominador $(b + d)$ é positivo, a positividade de Q^* requer que o numerador $(ad - bc)$ também seja positivo. Por conseguinte, para ser significativo em termos econômicos, o presente modelo deve conter a restrição adicional $ad > bc$.

O significado dessa restrição pode ser observado na Figura 3.1. Todos sabemos que P^* e Q^* de um modelo de mercado podem ser determinados graficamente na interseção das curvas de demanda e de oferta. Para que $Q^* > 0$, é preciso que o ponto de interseção esteja localizado acima do eixo horizontal na Figura 3.1, o que, por sua vez, requer que as inclinações e interseções das duas curvas com o eixo vertical obedeçam a uma certa restrição no que se refere a suas grandezas (magnitudes) relativas. Essa restrição, segundo (3.5), é $ad > bc$, dado que ambos, b e d , são positivos.

A propósito, a interseção das curvas de demanda e de oferta na Figura 3.1 não é conceitualmente diferente da interseção mostrada no diagrama de Venn da Figura 2.2b. Há apenas uma diferença: em vez de os pontos estarem dentro de dois círculos, o presente caso envolve os pontos que estão em duas retas. Denotemos o conjunto de pontos nas curvas de demanda e de oferta por D e S respectivamente. Então, utilizando o símbolo Q ($= Q_d = Q_s$), os dois conjuntos e sua interseção podem ser escritos

$$D = \{(P, Q) \mid Q = a - bP\}$$

$$S = \{(P, Q) \mid Q = -c + dP\}$$

e

$$D \cap S = (P^*, Q^*)$$

Nessa circunstância, o conjunto interseção contém somente um único elemento, o par ordenado (P^*, Q^*) . O equilíbrio de mercado é único.

EXERCÍCIO 3.2

1. Dado o modelo de mercado

$$Q_d = Q_s$$

$$Q_d = 21 - 3P$$

$$Q_s = -4 + 8P$$

Calcule P^* e Q^* por (a) eliminação de variáveis e (b) usando as fórmulas (3.4) e (3.5). (Use frações em vez de decimais.)

2. Considerando as seguintes funções de demanda e oferta:

$$(a) Q_d = 51 - 3P$$

$$(b) Q_d = 30 - 2P$$

$$Q_s = 6P - 10$$

$$Q_s = -6 + 5P$$

calcule P^* e Q^* por eliminação de variáveis. (Use frações em vez de decimais.)

3. Segundo (3.5), para Q^* ser positivo, é necessário que a expressão $(ad - bc)$ tenha o mesmo sinal algébrico de $(b + d)$. Verifique se essa condição é realmente satisfeita nos modelos dos Problemas 1 e 2.
4. Se $(b + d) = 0$ no modelo linear de mercado, pode-se achar uma solução de equilíbrio usando (3.4) e (3.5)? Justifique sua resposta.
5. Se $(b + d) = 0$ no modelo linear de mercado, o que você pode concluir em relação às posições das curvas de demanda e oferta na Figura 3.1? E, então, o que pode concluir em relação à solução de equilíbrio?

3.3 Equilíbrio parcial de mercado – um modelo não-linear

Vamos substituir a demanda linear no modelo de mercado isolado por uma função quadrática de demanda, mantendo linear a função de oferta. Além disso, vamos usar coeficientes numéricos em vez de parâmetros. Então, pode surgir um modelo como o seguinte:

$$Q_d = Q_s$$

$$Q_d = 4 - P^2$$

$$Q_s = 4P - 1 \quad (3.6)$$

Como anteriormente, esse sistema de três equações pode ser reduzido a uma única equação por eliminação de variáveis (por substituição):

$$4 - P^2 = 4P - 1$$

ou

$$P^2 + 4P - 5 = 0 \quad (3.7)$$

Essa é uma equação quadrática porque a expressão à esquerda é uma função quadrática da variável P . Uma importante diferença entre uma equação quadrática e uma equação linear é que, em geral, a primeira resultará em dois valores de solução.

Equação quadrática versus função quadrática

Antes de discutir o método de solução, é preciso fazer uma clara distinção entre os dois termos *equação quadrática* e *função quadrática*. Como discutimos anteriormente, a expressão $P^2 + 4P - 5$ constitui uma *função* quadrática, digamos, $f(P)$. Por conseguinte, podemos escrever

$$f(P) = P^2 + 4P - 5 \quad (3.8)$$

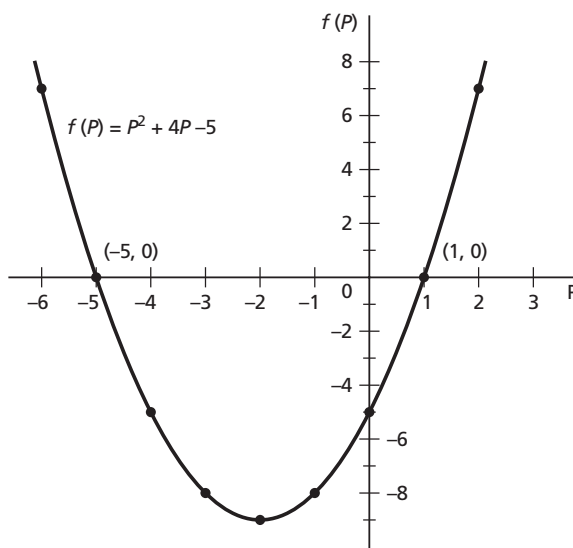
A expressão (3.8) especifica uma regra de mapeamento de P para $f(P)$, tal como

P	...	-6	-5	-4	-3	-2	-1	0	1	2	...
$f(P)$	...	7	0	-5	-8	-9	-8	-5	0	7	...

Embora tenhamos relacionado somente nove valores de P nessa tabela, na verdade, *todos* os valores de P no domínio da função podem figurar na lista. Talvez seja esta a razão por que raramente falamos em “resolver” a equação $f(P) = P^2 + 4P - 5$, pois normalmente esperamos que os “valores da solução” sejam poucos, mas aqui todos os valores de P podem ser envolvidos. Não obstante, podemos considerar legitimamente cada par ordenado na tabela – tal como $(-6, 7)$ e $(-5, 0)$ – como uma solução de (3.8), visto que cada um deles realmente satisfaz aquela equação. Considerando que pode-se escrever um número infinito desses pares ordenados, um para cada valor de P , há um número infinito de soluções para (3.8). Quando descritos por uma curva, o conjunto esses pares ordenados, juntos, resultam na parábola na Figura 3.2.

Em (3.7), onde determinamos que a função quadrática $f(P)$ é igual a zero, a situação muda fundamentalmente. Já que agora a variável $f(P)$ desaparece (pois recebeu um valor zero), o resultado é uma *equação* quadrática com uma única variável P .[†] Agora que $f(P)$ está restrita a um valor zero, somente um número selecionado de valores de P pode satisfazer (3.7) e se qualificar como seus valores da solução, a saber, os valores de P nos quais a parábola da Figura 3.2 intercepta o eixo horizontal – no qual $f(P)$ é zero. Note que, dessa vez, os valores da solução são apenas valores P , e não pares ordenados. Os valores da solução P normalmente são denominados as *raízes* da equação quadrática $f(P) = 0$ ou, alternativamente, os *zeros* da função quadrática $f(P)$.

FIGURA 3.2



[†] A distinção entre função quadrática e equação quadrática que acabamos de discutir também pode ser estendida para casos de outros polinômios, além dos quadráticos. Assim, resulta uma equação cúbica quando igualamos uma função cúbica a zero.

Há dois desses pontos de interseção na Figura 3.2, a saber, $(1, 0)$ e $(-5, 0)$. Como exigido, o segundo elemento de cada um desses pares ordenados (a *ordenada* do ponto correspondente) mostra que $f(P) = 0$ em ambos os casos. Por outro lado, o primeiro elemento de cada par ordenado (a *abscissa* do ponto) dá o valor da solução de P . Aqui, obtemos duas soluções:

$$P_1^* = 1 \quad \text{e} \quad P_2^* = -5$$

mas somente a primeira é admissível em termos econômicos, pois a regra exclui preços negativos.

A fórmula quadrática

A equação (3.7) foi resolvida graficamente, mas também há um método algébrico disponível. Em geral, dada uma equação quadrática na forma

$$ax^2 + bx + c = 0 \quad (a \neq 0) \quad (3.9)$$

há duas raízes que podem ser obtidas da *fórmula quadrática*:

$$x_1^*, x_2^* = \frac{-b \pm (b^2 - 4ac)^{1/2}}{2a} \quad (3.10)$$

onde a parte referente ao $+$ no sinal $\pm$ resulta em x_1^* e parte referente ao $-$ resulta em x_2^* .

Note também que, contanto que $b^2 - 4ac > 0$, os valores de x_1^* e x_2^* serão diferentes, o que nos dá dois números reais distintos como raízes. Mas, no caso especial em que $b^2 - 4ac = 0$, veríamos que $x_1^* = x_2^* = -b/2a$. Nesse caso, as duas raízes compartilham o valor idêntico; são denominadas *raízes repetidas*. Em mais um outro caso especial, onde $b^2 - 4ac < 0$, teríamos de extrair a raiz quadrada de um número negativo, o que não é possível no sistema dos números reais. Nesse último caso, não existe nenhuma raiz de valor real. Discutiremos mais esse assunto na Seção 16.1.

Essa fórmula amplamente utilizada é obtida por meio de um processo conhecido como “completar o quadrado”. Em primeiro lugar, dividir cada termo de (3.9) por a resulta na equação

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

Subtraindo c/a e adicionando $b^2/4a^2$ a ambos os lados da equação, obtemos

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

O lado esquerdo agora é um “quadrado perfeito” e, assim, a equação pode ser expressa como

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

ou, após extrair a raiz quadrada de ambos os lados,

$$x + \frac{b}{2a} = \pm \frac{(b^2 - 4ac)^{1/2}}{2a}$$

Finalmente, subtraindo $b/2a$ de ambos os lados, obtém-se o resultado em (3.10).

Aplicando a fórmula a (3.7), onde $a = 1$, $b = 4$, $c = -5$ e $x = P$, verificamos que as raízes são

$$P_1^*, P_2^* = \frac{-4 \pm (16 + 20)^{1/2}}{2} = \frac{-4 \pm 6}{2} = 1, -5$$

que estão de acordo com as soluções gráficas na Figura 3.2. Novamente, rejeitamos $P_2^* = -5$ com base em premissas econômicas e, após omitir o subscrito 1, escrevemos simplesmente $P^* = 1$.

Com essa informação em mãos, a quantidade de equilíbrio Q^* pode ser calculada prontamente pela segunda ou pela terceira equação de (3.6), ou seja, $Q^* = 3$.

Uma outra solução gráfica

Um dos métodos de solução gráfica do presente modelo foi apresentado na Figura 3.2. Contudo, uma vez que a variável quantidade foi eliminada na obtenção da equação quadrática, somente P^* pode ser calculado a partir daquela figura. Se estivermos interessados em achar P^* e Q^* simultaneamente por um gráfico, devemos usar um diagrama que apresente Q em um eixo e P no outro, semelhante ao da Figura 3.1. Isso é ilustrado na Figura 3.3. É claro que, novamente, nosso problema é achar a interseção de dois conjuntos de pontos, a saber,

$$D = \{(P, Q) \mid Q = 4 - P^2\}$$

e

$$S = \{(P, Q) \mid Q = 4P - 1\}$$

Se não for imposta nenhuma restrição sobre o domínio e a imagem, o conjunto interseção conterá dois elementos, a saber,

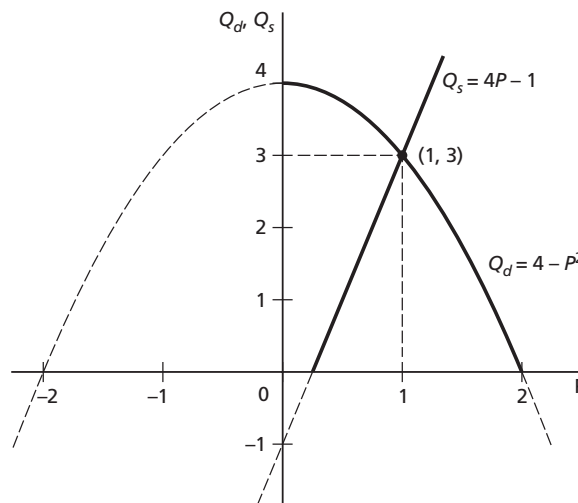
$$D \cap S = \{(1, 3), (-5, -21)\}$$

O primeiro está localizado no quadrante I, e o último (não desenhado) no quadrante III. Se, por restrição, o domínio e a imagem tiverem de ser não-negativos, somente o primeiro par ordenado $(1, 3)$ pode ser aceito. Então, mais uma vez, o equilíbrio é único.

Equações polinomiais de grau mais alto

Se um sistema de equações simultâneas não for redutível a uma equação linear como (3.3)[†] ou a uma equação quadrática como (3.7), mas a uma equação cúbica (polinomial de terceiro grau) ou quártica (polinomial de quarto grau), será mais difícil achar as raízes. Um método útil, que pode funcionar, é a *fatoração* da função.

FIGURA 3.3



[†] A equação (3.3) pode ser considerada como o resultado de igualar a função linear $(b + d)P - (a + c)$ a zero.

EXEMPLO 1

A expressão $x^3 - x^2 - 4x + 4$ pode ser escrita como o produto de três fatores $(x - 1)$, $(x + 2)$ e $(x - 2)$. Assim, a equação cúbica

$$x^3 - x^2 - 4x + 4 = 0$$

pode ser escrita, após a fatoração, como

$$(x - 1)(x + 2)(x - 2) = 0$$

Para que o produto à esquerda seja zero, pelo menos um dos três termos do produto deve ser zero. Igualando a zero um termo de cada vez, obtemos

$$x - 1 = 0 \quad \text{ou} \quad x + 2 = 0 \quad \text{ou} \quad x - 2 = 0$$

Essas três equações fornecerão as três raízes da equação cúbica, a saber,

$$x_1^* = 1 \quad x_2^* = -2 \quad \text{e} \quad x_3^* = 2$$

O Exemplo 1 ilustra dois fatos interessantes e úteis sobre fatoração. Primeiro, dada uma equação polinomial de terceiro grau, a fatoração resulta em três termos na forma $(x - \text{raiz})$, resultando, assim, em três raízes. Em geral, uma equação polinomial de grau n deve ter um total de n raízes. Segundo, e mais importante para a finalidade de achar raízes, observamos a seguinte relação entre as três raízes $(1, -2, 2)$ e o termo constante 4: uma vez que o termo constante deve ser o produto das três raízes, cada raiz deve ser um divisor do termo constante. Essa relação pode ser formalizada no seguinte teorema:

Teorema I Dada a equação polinomial

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

na qual todos os coeficientes são inteiros e o coeficiente de x^n é a unidade, se existirem raízes inteiras, então cada uma delas deve ser um divisor de a_0 .

Entretanto, às vezes encontramos coeficientes fracionários na equação polinomial, como em

$$x^4 + \frac{5}{2}x^3 - \frac{11}{2}x^2 - 10x + 6 = 0$$

que não cumprem a condição do Teorema I. Mesmo multiplicando todos os termos por 2 para eliminar as frações (o que resulta na forma mostrada no Exemplo 2 a seguir), ainda assim não podemos aplicar o Teorema I porque o coeficiente do termo de grau mais alto não é a unidade. Nesses casos, podemos recorrer a um teorema mais geral:

Teorema II Dada a equação polinomial com coeficientes inteiros

$$a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

se existir uma raiz racional r/s , onde r e s são inteiros que não têm um divisor comum, exceto a unidade, então r é um divisor de a_0 e s é um divisor de a_n .

EXEMPLO 2

A equação quártica

$$2x^4 + 5x^3 - 11x^2 - 20x + 12 = 0$$

tem raízes racionais? Com $a_0 = 12$, os únicos valores possíveis para o numerador r em r/s pertencem ao conjunto de divisores $\{1, -1, 2, -2, 3, -3, 4, -4, 6, -6, 12, -12\}$. E, para $a_n = 2$, os únicos va-

lores possíveis para r/s pertencem ao conjunto de divisores $\{1, -1, 2, -2\}$. Tomando cada elemento no conjunto r de cada vez, e dividindo-o por elemento no conjunto s , respectivamente, verificamos que r/s somente pode assumir os valores

$$1, -1, \frac{1}{2}, -\frac{1}{2}, 2, -2, 3, -3, \frac{3}{2}, -\frac{3}{2}, 4, -4, 6, -6, 12, -12$$

Entre esses candidatos a raízes, muitos não satisfazem a equação dada. Por exemplo, se $x = 1$ na equação quártica obtemos o ridículo resultado $-12 = 0$. Na verdade, como estamos resolvendo uma equação quártica, podemos esperar que no máximo quatro dos valores r/s relacionados anteriormente se qualifiquem como raízes. Os quatro candidatos bem-sucedidos são: $\frac{1}{2}$, 2 , -2 e -3 . Assim, conforme o princípio da fatoração, podemos escrever a equação quártica equivalente como

$$(x - \frac{1}{2})(x - 2)(x + 2)(x + 3) = 0$$

onde o primeiro fator também pode ser escrito alternativamente como $(2x - 1)$.

No Exemplo 2, rejeitamos o candidato a raiz 1 porque $x = 1$ não satisfaz a equação dada, isto é, substituir $x = 1$ na equação não produz a identidade $0 = 0$ como requerido. Agora, considere o caso em que $x = 1$ é, de fato, uma raiz de alguma equação polinomial. Nesse caso, visto que $x^n = x^{n-1} = \dots = x = 1$, a equação polinomial se reduziria à forma simples $a_n + a_{n-1} + \dots + a_1 + a_0 = 0$. Esse fato fornece o princípio racional para o seguinte teorema:

Teorema III Dada a equação polinomial

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$$

se os coeficientes $a_n, a_{n-1}, \dots, a_0$ somarem zero, então $x = 1$ é uma raiz da equação.

EXERCÍCIO 3.3

- Encontre os zeros das seguintes funções, por meio de um gráfico:
 - $f(x) = x^2 - 8x + 15$
 - $g(x) = 2x^2 - 4x - 16$
- Resolva o Problema 1 pela fórmula quadrática.
- Ache uma equação cúbica com raízes 6, -1 e 3.
 - Ache uma equação quártica com raízes 1, 2, 3 e 5.
- Para cada uma das seguintes equações polinomiais, determine se $x = 1$ é uma raiz.
 - $x^3 - 2x^2 + 3x - 2 = 0$
 - $3x^4 - x^2 + 2x - 4 = 0$
 - $2x^3 - \frac{1}{2}x^2 + x - 2 = 0$
- Calcule as raízes racionais, se existirem, das seguintes equações:
 - $x^3 - 4x^2 + x + 6 = 0$
 - $x^3 + \frac{3}{4}x^2 - \frac{3}{8}x - \frac{1}{8} = 0$
 - $8x^3 + 6x^2 - 3x - 1 = 0$
 - $x^4 - 6x^3 + 7\frac{3}{4}x^2 - \frac{3}{2}x - 2 = 0$
- Encontre a solução de equilíbrio para cada um dos seguintes modelos:
 - $Q_d = Q_s$
 $Q_d = 3 - P^2$
 $Q_s = 6P - 4$
 - $Q_d = Q_s$
 $Q_d = 8 - P^2$
 $Q_s = P^2 - 2$
- A condição de equilíbrio de mercado, $Q_d = Q_s$, é freqüentemente expressa em uma forma alternativa equivalente, $Q_d - Q_s = 0$, cuja interpretação econômica é a seguinte: "o excesso de demanda é zero". A equação (3.7) representa essa última versão da condição de equilíbrio? Se a resposta for negativa, dê uma interpretação econômica adequada para (3.7).

3.4 Equilíbrio geral de mercado

As duas últimas seções trataram de modelos de um mercado isolado, no qual Q_d e Q_s de uma mercadoria são funções do preço daquela mercadoria apenas. No mundo real, entretanto, nenhuma mercadoria jamais goza (ou sofre) tal existência eremítica; para cada mercadoria, normalmente existiriam muitos bens substitutos e complementares. Assim, uma descrição mais realista da função de demanda de uma mercadoria deve levar em conta não apenas o efeito do preço da mercadoria em si, mas também os preços de mercadorias relacionadas. O mesmo vale também para a função de oferta. Tão logo os preços de outras mercadorias passem a fazer parte do quadro, entretanto, a própria estrutura do modelo tem de ser ampliada, de modo a poder resultar também os valores de equilíbrio desses outros preços. Conseqüentemente, as variáveis preço e quantidade de várias mercadorias devem entrar no modelo endogenamente, em massa.

Em um modelo de mercado isolado, a condição de equilíbrio consiste em apenas uma equação, $Q_d = Q_s$, ou $E \equiv Q_d - Q_s = 0$, onde E representa o excesso de demanda. Quando várias mercadorias são consideradas simultaneamente, o equilíbrio exigiria a ausência de excesso de demanda para cada uma das mercadorias incluídas no modelo, pois, se ao menos *uma* mercadoria enfrentar um excesso de demanda, o ajuste de preço dessa mercadoria necessariamente afetará as quantidades demandadas e ofertadas das mercadorias relacionadas, ocasionando, assim, uma mudança geral de preços. Conseqüentemente, a condição de equilíbrio para um modelo de mercado de n mercadorias envolveria n equações, uma para cada mercadoria, na forma

$$E_i \equiv Q_{di} - Q_{si} = 0 \quad (i = 1, 2, \dots, n) \quad (3.11)$$

Se existir uma solução, haverá um conjunto de preços P_i^* e quantidades correspondentes Q_i^* tais que todas as n equações da condição de equilíbrio serão satisfeitas simultaneamente.

Modelo de mercado de duas mercadorias

Para ilustrar esse problema, vamos discutir um modelo simples no qual somente duas mercadorias são relacionadas uma com a outra. Por simplicidade, admitimos que as funções de demanda e de oferta de ambas as mercadorias são lineares. Em termos paramétricos, tal modelo pode ser escrito como

$$\begin{aligned} Q_{d1} - Q_{s1} &= 0 \\ Q_{d1} &= a_0 + a_1 P_1 + a_2 P_2 \\ Q_{s1} &= b_0 + b_1 P_1 + b_2 P_2 \\ Q_{d2} - Q_{s2} &= 0 \\ Q_{d2} &= \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 \\ Q_{s2} &= \beta_0 + \beta_1 P_1 + \beta_2 P_2 \end{aligned} \quad (3.12)$$

onde os coeficientes a e b pertencem às funções de demanda e de oferta da primeira mercadoria e os coeficientes α e β são atribuídos à demanda e oferta da segunda. Não nos preocupamos em especificar os sinais dos coeficientes, mas, no decorrer da análise, surgirão certas restrições como pré-requisito para resultados econômicos sensatos. Além disso, em um exemplo numérico subseqüente, serão feitos alguns comentários sobre os sinais específicos que devem ser atribuídos aos coeficientes.

Como uma primeira etapa da solução desse modelo, podemos recorrer novamente à eliminação de variáveis. Substituindo a segunda e a terceira equações na primeira (para a primeira mercadoria) e a quinta e a sexta equações na quarta (para a segunda mercadoria), o modelo é reduzido a duas equações de duas variáveis:

$$(a_0 - b_0) + (a_1 - b_1)P_1 + (a_2 - b_2)P_2 = 0$$

$$(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)P_1 + (\alpha_2 - \beta_2)P_2 = 0 \quad (3.13)$$

Essas duas equações representam a versão de (3.11) para duas mercadorias, após a substituição das funções de demanda e de oferta nas duas condições de equilíbrio.

Embora esse seja um sistema simples de somente duas equações, há 12 parâmetros envolvidos, e será impraticável recorrer a manipulações algébricas a menos que seja utilizado algum tipo de notação abreviada. Portanto, vamos definir os símbolos abreviados

$$\begin{aligned} c_i &\equiv a_i - b_i \\ \gamma_i &\equiv \alpha_i - \beta_i \end{aligned} \quad (i = 0, 1, 2)$$

Então, após transportar os termos c_0 e γ_0 para o lado direito, obtemos

$$\begin{aligned} c_1P_1 + c_2P_2 &= -c_0 \\ \gamma_1P_1 + \gamma_2P_2 &= -\gamma_0 \end{aligned} \quad (3.13')$$

que podem ser resolvidas por mais eliminação de variáveis. Na primeira equação, pode-se achar $P_2 = -(c_0 + c_1P_1)/c_2$. Substituindo essa expressão na segunda equação e resolvendo, obtemos

$$P_1^* = \frac{c_2\gamma_0 - c_0\gamma_2}{c_1\gamma_2 - c_2\gamma_1} \quad (3.14)$$

Note que P_1^* está expresso integralmente em termos dos dados (parâmetros) do modelo, como deve ser expresso um valor da solução. Utilizando um processo semelhante, achamos o preço de equilíbrio da segunda mercadoria:

$$P_2^* = \frac{c_0\gamma_1 - c_1\gamma_0}{c_1\gamma_2 - c_2\gamma_1} \quad (3.15)$$

Contudo, para que esses dois valores façam sentido, devem ser impostas certas restrições ao modelo. Em primeiro lugar, visto que a divisão por zero é indefinida, devemos exigir que o denominador comum de (3.14) e (3.15) seja não-zero (diferente de zero), isto é, $c_1\gamma_2 \neq c_2\gamma_1$. Em segundo lugar, para garantir a positividade, o numerador também deve ter o mesmo sinal que o denominador.

Encontrados os preços de equilíbrio, as quantidades de equilíbrio Q_1^* e Q_2^* podem ser prontamente calculadas substituindo (3.14) e (3.15) na segunda (ou na terceira) equação e na quinta (ou na sexta) equação de (3.12). Esses valores da solução também serão naturalmente expressos em termos de parâmetros. (O cálculo propriamente dito desses valores é proposto como um exercício.)

Exemplo numérico

Suponha que as funções de demanda e oferta sejam as seguintes, em termos numéricos:

$$\begin{aligned} Q_{d1} &= 10 - 2P_1 + P_2 \\ Q_{s1} &= -2 + 3P_1 \\ Q_{d2} &= 15 + P_1 - P_2 \\ Q_{s2} &= -1 + 2P_2 \end{aligned} \quad (3.16)$$

Qual é a solução de equilíbrio?

Antes de responder a essa pergunta, vamos examinar os coeficientes numéricos. Para cada mercadoria, verifica-se que Q_{si} depende apenas de P_i , mas que Q_{di} é uma função de ambos os preços. Note que, enquanto P_1 tem um coeficiente negativo em Q_{d1} , como seria de se esperar, o coeficiente de P_2 é positivo. O fato de que um aumento em P_2 tende a elevar Q_{d1} sugere que as duas mercadorias são substitutas uma da outra. O papel de P_1 na função Q_{d2} tem uma interpretação semelhante.

Com esses coeficientes, os símbolos abreviados c_i e γ_i assumirão os seguintes valores:

$$\begin{aligned} c_0 &= 10 - (-2) = 12 & c_1 &= -2 - 3 = -5 & c_2 &= 1 - 0 = 1 \\ \gamma_0 &= 15 - (-1) = 16 & \gamma_1 &= 1 - 0 = 1 & \gamma_2 &= -1 - 2 = -3 \end{aligned}$$

Substituindo esses valores diretamente em (3.14) e (3.15), obtemos

$$P_1^* = \frac{52}{14} = 3\frac{5}{7} \quad \text{e} \quad P_2^* = \frac{92}{14} = 6\frac{4}{7}$$

E a substituição subsequente de P_1^* e P_2^* em (3.16) resulta em

$$Q_1^* = \frac{64}{7} = 9\frac{1}{7} \quad \text{e} \quad Q_2^* = \frac{85}{7} = 12\frac{1}{7}$$

Assim, todos os valores de equilíbrio resultaram positivos, como requerido. Para preservar os valores exatos de P_1^* e P_2^* que serão usados no cálculo subsequente de Q_1^* e Q_2^* , é aconselhável expressá-los em frações em vez de decimais.

Poderíamos ter obtido os preços de equilíbrio graficamente? A resposta é sim. Examinando (3.13), fica claro que um modelo de duas mercadorias pode ser resumido por duas equações de duas variáveis, P_1 e P_2 . Com coeficientes numéricos conhecidos, ambas as equações podem ser colocadas no plano cartesiano P_1P_2 e, então, a interseção das duas curvas indicará P_1^* e P_2^* .

Caso de n mercadorias

A discussão anterior do mercado multimercadorias limitou-se ao caso de duas mercadorias, mas já deve ser evidente que estamos passando da análise de *equilíbrio parcial* para a análise de *equilíbrio geral*. À medida que mais mercadorias entram em um modelo, haverá mais variáveis e mais equações, e as equações ficarão mais compridas e mais complicadas. Se todas as mercadorias de uma economia fossem incluídas em um modelo de mercado abrangente, o resultado seria um modelo de equilíbrio geral de mercado do tipo Walrasiano, no qual o excesso de demanda para cada mercadoria é considerado uma função dos preços de todas as mercadorias na economia.

É claro que alguns desses preços podem ter coeficientes zero quando não desempenham nenhum papel na determinação do excesso de demanda de uma mercadoria específica; por exemplo, na função de excesso de demanda de pianos, o preço da pipoca pode perfeitamente ter um coeficiente zero. Contudo, em geral, com n mercadorias ao todo, podemos expressar as funções de demanda e de oferta como se segue (usando Q_{di} e Q_{si} como os símbolos de funções no lugar de f e g):

$$\begin{aligned} Q_{di} &= Q_{di}(P_1, P_2, \dots, P_n) \\ Q_{si} &= Q_{si}(P_1, P_2, \dots, P_n) \end{aligned} \quad (i = 1, 2, \dots, n) \quad (3.17)$$

Em vista do índice subscrito, essas duas equações representam a totalidade das $2n$ funções que o modelo contém. (Essas funções não são necessariamente lineares.) Além disso, condição de equilíbrio em si é composta por um conjunto de n equações,

$$Q_{di} - Q_{si} = 0 \quad (i = 1, 2, \dots, n) \quad (3.18)$$

Quando (3.18) é adicionada a (3.17), o modelo fica completo. Conseqüentemente, a contagem final é um total de $3n$ equações.

Com a substituição de (3.17) em (3.18), contudo, o modelo pode ser reduzido a um conjunto de n equações simultâneas:

$$Q_{di}(P_1, P_2, \dots, P_n) - Q_{si}(P_1, P_2, \dots, P_n) = 0 \quad (i = 1, 2, \dots, n)$$

Além disso, considerando que $E_i \equiv Q_{di} - Q_{si}$, onde E_i também é, necessariamente, uma função de todos os n preços, o último conjunto de equações pode ser escrito alternativamente como

$$E_i(P_1, P_2, \dots, P_n) = 0 \quad (i = 1, 2, \dots, n)$$

Resolvidas simultaneamente, essas n equações podem determinar os n preços de equilíbrio P_i^* – se realmente existir uma solução. E então, Q_i^* pode ser obtido das funções de demanda ou de oferta.

Solução de um sistema geral de equações

Se um modelo vier equipado com coeficientes numéricos, como em (3.16), os valores de equilíbrio das variáveis também serão termos numéricos. Em um nível mais geral, se um modelo for expresso em termos de constantes paramétricas, como em (3.12), os valores de equilíbrio também envolverão parâmetros e, conseqüentemente, aparecerão como “fórmulas”, como exemplificado por (3.14) e (3.15). Entretanto, se, para maior generalidade, até mesmo as formas de funções deixarem de ser especificadas em um modelo, como em (3.17), a maneira de expressar os valores da solução também será, inevitavelmente, excessivamente geral.

Lançando mão de nossa experiência em modelos paramétricos, sabemos que um valor da solução é sempre uma expressão em termos dos parâmetros. Para um modelo de função geral que contenha, digamos, um total de m parâmetros $(a_1, a_2, \dots, a_m)$ – onde m não é necessariamente igual a n – podemos esperar que os n preços de equilíbrio tomem a forma analítica geral de

$$P_i^* = P_i^*(a_1, a_2, \dots, a_m) \quad (i = 1, 2, \dots, n) \quad (3.19)$$

Essa é uma afirmação simbólica no sentido de que o valor da solução de *cada* variável (aqui, o preço) é uma função do conjunto de todos os parâmetros do modelo. Como essa observação é muito geral, na verdade ela não dá informações muito detalhadas sobre a solução. Mas, no tratamento analítico geral de alguns tipos de problema, mesmo esse modo aparentemente não-informativo de expressar um solução provará ser útil, como veremos no Capítulo 8.

Escrever uma solução desse tipo é uma tarefa fácil. Mas existe uma desvantagem: a expressão em (3.19) pode ser justificada se, e somente se, existir, de fato, uma solução *única*, pois então, e somente então, podemos mapear os m parâmetros ordenados $(a_1, a_2, \dots, a_m)$ para um determinado valor para cada preço P_i^* . Ainda assim, infelizmente para nós, não há nenhuma razão *a priori* para supor que cada modelo automaticamente resultará em uma solução única. Com relação a isso, é preciso enfatizar que o processo de “contar equações e incógnitas” não é um teste suficiente. Alguns exemplos simples deverão nos convencer de que um número igual de equações e incógnitas (variáveis endógenas) não garantem, necessariamente, a existência de uma solução única.

Considere os três sistemas de equações simultâneas

$$\begin{aligned} x + y &= 8 \\ x + y &= 9 \end{aligned} \quad (3.20)$$

$$\begin{aligned} 2x + y &= 12 \\ 4x + 2y &= 24 \end{aligned} \quad (3.21)$$

$$\begin{aligned} 2x + 3y &= 58 \\ y &= 18 \\ x + y &= 20 \end{aligned} \quad (3.22)$$

Em (3.20), a despeito do fato de que duas incógnitas estão ligadas por exatamente duas equações, ainda assim não existe nenhuma solução. Acontece que essas equações são *inconsistentes* pois, se a soma de x e y é 8, então não pode ser 9 ao mesmo tempo. Em (3.21), um outro caso de duas equações com duas variáveis, as duas equações são *funcionalmente dependentes*, o que significa que uma pode ser derivada da outra e está implícita na outra. (Aqui, a segunda equação é igual a duas vezes a primeira equação.) Conseqüentemente, uma das equações é redundante e pode ser descartada do sistema, deixando, efetivamente, somente uma equação com duas incógnitas. Então, a solução será a equação $y = 12 - 2x$, que não resulta em um único par ordenado (x^*, y^*) , mas em um número infinito de pares, incluindo $(0, 12)$, $(1, 10)$, $(2, 8)$ etc., todos os quais satisfazem aquela equação. Por fim, o caso de (3.22) envolve mais equações que incógnitas; ainda assim, o par ordenado $(2, 18)$ constitui sua única solução. A razão é que, em vista da existência de dependência funcional entre as equações (a primeira é igual à segunda mais duas vezes a terceira), temos, na verdade, somente duas equações independentes, consistentes, com duas variáveis.

Esses exemplos simples devem ser suficientes para transmitir a importância da *consistência* e da *independência funcional* como os dois pré-requisitos para a aplicação do processo de contagem de equações e incógnitas. Em geral, para aplicar esse processo, certifique-se de que (1) a satisfação de qualquer uma das equações no modelo não impede a satisfação de uma outra; e (2) nenhuma equação é redundante. Em (3.17), por exemplo, podemos admitir com segurança que as n funções de demanda e as n funções de oferta são independentes umas das outras, pois cada uma é obtida de uma fonte diferente – cada demanda das decisões de um grupo de consumidores e cada oferta da decisão de um grupo de empresas. Assim, cada função serve para descrever uma faceta da situação do mercado e nenhuma é redundante. Talvez também seja possível admitir a consistência mútua. Além disso, as equações de condição de equilíbrio em (3.18) também são independentes e presumivelmente consistentes. Por conseguinte, a solução analítica como escrita em (3.19) pode, em geral, ser considerada justificável.[†]

Para modelos de equações simultâneas, há métodos sistemáticos para testar a existência de uma única (ou determinada) solução. Esses testes envolveriam, para modelos lineares, uma aplicação do conceito de *determinantes*, que será apresentado no Capítulo 5. No caso de modelos não-lineares, tal teste exigiria também o conhecimento de derivadas parciais e de um tipo especial de determinante chamado *determinante Jacobiano*, que será discutido nos Capítulos 7 e 8.

EXERCÍCIO 3.4

1. Estude, etapa a etapa, a solução de (3.13'), verificando, assim, os resultados em (3.14) e (3.15).
2. Escreva (3.14) e (3.15) em termos dos parâmetros originais do modelo em (3.12).
3. As funções de demanda e de oferta de um modelo de mercado de duas mercadorias são as seguintes:

$$\begin{aligned} Q_{d1} &= 18 - 3P_1 + P_2 & Q_{d2} &= 12 + P_1 - 2P_2 \\ Q_{s1} &= -2 + 4P_1 & Q_{s2} &= -2 + 3P_2 \end{aligned}$$

Calcule P_i^* e Q_i^* ($i = 1, 2$). (Use frações em vez de decimais.)

[†] Esse é, em essência, o modo como Léon Walras abordou o problema da existência de um equilíbrio geral de mercado. Na literatura moderna, podemos encontrar inúmeras provas matemáticas sofisticadas da existência de um equilíbrio de mercado competitivo sob certas condições econômicas postuladas. Porém, nesse caso, é utilizada matemática avançada. A prova mais fácil de entender seria, talvez, a dada por Dorfman, Robert; Samuelson, Paul A.; e M. Robert; Nova York: Solow. In: *Linear Programming and Economic Analysis*. McGraw-Hill Book Company, 1958, Capítulo 13.

3.5 Equilíbrio na análise da renda nacional

Embora até aqui a discussão da análise estática tenha se restringido a *modelos de mercado* sob vários aspectos – lineares e não-lineares, de uma mercadoria e multimercomodias, específicos e gerais – é claro que essa análise também pode ser aplicada em outras áreas da economia. Como exemplo, podemos citar o mais simples dos modelos: o keynesiano de renda nacional,

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= a + bY \end{aligned} \quad (a > 0, \quad 0 < b < 1) \quad (3.23)$$

onde Y e C representam as variáveis endógenas renda nacional e despesa de consumo (planejada), respectivamente, e I_0 e G_0 representam despesas de investimentos e governamentais determinadas exogenamente. A primeira equação é uma condição de equilíbrio (renda nacional = despesa total planejada). A segunda equação, a função de consumo, é comportamental. Os dois parâmetros da função de consumo, a e b , representam a despesa de consumo autônoma e a propensão marginal ao consumo, respectivamente.

É bem claro que essas duas equações de duas variáveis endógenas não dependem funcionalmente uma da outra, nem são inconsistentes uma em relação à outra. Assim, poderíamos achar os valores de equilíbrio da renda e da despesa de consumo, Y^* e C^* , em termos dos parâmetros a e b e das variáveis exógenas I_0 e G_0 .

A substituição da segunda equação na primeira reduzirá (3.23) a uma única equação de uma variável, Y :

$$Y = a + bY + I_0 + G_0$$

ou $(1 - b)Y = a + I_0 + G_0$ (reunindo termos que envolvem Y)

Para encontrar o valor da solução de Y (renda nacional de equilíbrio), basta dividir tudo por $(1 - b)$:

$$Y^* = \frac{a + I_0 + G_0}{1 - b} \quad (3.24)$$

Note, mais uma vez, que o valor da solução é expresso integralmente em termos dos parâmetros e variáveis exógenas, os dados determinados pelo modelo. Inserir (3.24) na segunda equação de (3.23) resultará, então, no nível de equilíbrio de despesa de consumo:

$$\begin{aligned} C^* &= a + bY^* = a + \frac{b(a + I_0 + G_0)}{1 - b} \\ &= \frac{a(1 - b) + b(a + I_0 + G_0)}{1 - b} = \frac{a + b(I_0 + G_0)}{1 - b} \end{aligned} \quad (3.25)$$

Novamente, isso é expresso inteiramente em termos dos dados determinados pelo modelo.

Ambos, Y^* e C^* , têm a expressão $(1 - b)$ no denominador; daí a necessidade de impor uma restrição $b \neq 1$, para evitar divisão por zero. Visto que admitimos que b , a propensão marginal ao consumo, é uma fração positiva, essa restrição é automaticamente satisfeita. Além do mais, para que Y^* e C^* sejam positivos, os numeradores em (3.24) e (3.25) devem ser positivos. Como as despesas exógenas I_0 e G_0 normalmente são positivas, assim como o parâmetro a (a interseção vertical da função de consumo), o sinal das expressões do numerador também funcionará.

Para verificar nosso cálculo, podemos adicionar a expressão C^* em (3.25) a $(I_0 + G_0)$ e verificar se a soma é igual à expressão Y^* em (3.24).

Esse é um modelo cuja crueza e simplicidade extremas são óbvias, mas outros modelos de determinação da renda nacional com graus variáveis de complexidade e sofisticação também podem ser construídos. Contudo, em cada caso, os princípios envolvidos na construção e na análise do modelo são idênticos aos já discutidos. Por essa razão, não iremos mais adiante com nossas ilustrações. Um modelo de renda nacional mais abrangente, envolvendo o equilíbrio simultâneo do mercado monetário e do mercado de bens será discutido na Seção 8.6.

EXERCÍCIO 3.5

1. Dado o seguinte modelo:

$$Y = C + I_0 + G_0$$

$$C = a + b(Y - T) \quad (a > 0, \quad 0 < b < 1) \quad [T: \text{impostos}]$$

$$T = d + tY \quad (d > 0, \quad 0 < t < 1) \quad [t: \text{taxa de imposto sobre a renda}]$$

(a) Quantas variáveis endógenas existem?

(b) Calcule Y^* , T^* , e C^* .

2. Seja o modelo de renda nacional:

$$Y = C + I_0 + G$$

$$C = a + b(Y - T_0) \quad (a > 0, \quad 0 < b < 1)$$

$$G = gY \quad (0 < g < 1)$$

(a) Identifique as variáveis endógenas.

(b) Dê o significado econômico do parâmetro g .

(c) Calcule a renda nacional de equilíbrio.

(d) Quais restrições devem ser impostas aos parâmetros para que exista uma solução?

3. Calcule Y^* e C^* do seguinte:

$$Y = C + I_0 + G_0$$

$$C = 25 + 6Y^{1/2}$$

$$I_0 = 16$$

$$G_0 = 14$$

CAPÍTULO 4

Modelos lineares e álgebra matricial

Para o modelo de uma mercadoria (3.1), as soluções P^* e Q^* expressas em (3.4) e (3.5), respectivamente, são relativamente simples, mesmo havendo inúmeros parâmetros envolvidos. À medida que mais e mais mercadorias são incorporadas ao modelo, essas fórmulas de solução tornam-se rapidamente incômodas e impraticáveis. É por isso que tivemos de recorrer a uma certa notação abreviada, mesmo no caso de duas mercadorias – de modo que as soluções em (3.14) e (3.15) ainda possam ser escritas de maneira relativamente concisa. Não tentamos discutir nenhum modelo de três ou quatro mercadorias, mesmo na versão linear, primordialmente porque ainda não tínhamos à nossa disposição um método adequado para resolver um grande sistema de equações simultâneas. Esse método é encontrado na *álgebra matricial*, assunto deste e do próximo capítulo.

A álgebra matricial pode nos capacitar a fazer muitas coisas. Em primeiro lugar, proporciona um modo compacto de escrever um sistema de equações, mesmo que ele seja muito grande. Em segundo lugar, leva a um modo de testar a existência de uma solução pela avaliação de um *determinante* – um conceito que se relaciona muito de perto com o de uma matriz. Em terceiro lugar, nos dá um método para achar aquela solução (se ela existir). Uma vez que há sistemas de equações não somente na análise estática, mas também nas análises comparativas estática e dinâmica e nos problemas de otimização, o leitor encontrará ampla aplicação de álgebra matricial em quase todos os capítulos que vêm a seguir. Por essa razão, é bom introduzir a álgebra matricial logo de início.

Entretanto, a álgebra matricial apresenta uma pequena desvantagem: só pode ser aplicada a sistemas de equações *lineares*. E o grau de realismo com que equações lineares podem descrever relações econômicas propriamente ditas depende, é claro, da natureza das relações em questão. Em muitos casos, mesmo que admitir a linearidade implique um certo sacrifício do realismo, uma relação admitida como linear pode produzir uma aproximação suficientemente boa com uma relação não-linear para que sua utilização seja viável.

Em outros casos, embora preservando a não-linearidade do modelo, podemos efetuar uma transformação de variáveis de modo a obter uma relação linear com a qual trabalhar. Por exemplo, tomando o logaritmo de ambos os lados da função não-linear

$$y = ax^b$$

obtemos imediatamente a função

$$\log y = \log a + b \log x$$

O modelo de mercado de duas mercadorias (3.12) pode ser escrito – após a eliminação das variáveis de quantidade – como um sistema de duas equações lineares na forma (3.13'),

$$\gamma_1 P_1 + \gamma_2 P_2 = -\gamma_0$$

onde os parâmetros c_0 e γ_0 aparecem do lado direito do sinal de igualdade. Em geral, um sistema de m equações lineares com n variáveis $(x_1, x_2, \dots, x_n)$ também pode ser disposto no seguinte formato:

Em (4.1), a variável x_1 aparece somente na coluna da extrema esquerda e, em geral, a variável x_j aparece somente na j -ésima coluna no lado esquerdo do sinal de igualdade. O símbolo de parâmetro com dois índices, a_{ij} , representa o coeficiente que aparece na i -ésima equação e ligado à j -ésima variável. Por exemplo, a_{21} é o coeficiente da segunda equação ligado à variável x_1 . Por outro lado, o parâmetro d_i , que não está ligado a nenhuma variável, representa o termo constante na i -ésima equação. Por exemplo, d_1 é o termo constante na primeira equação. Consequentemente, todos os índices estão de acordo com as localizações específicas das variáveis e parâmetros em (4.1).

Há, essencialmente, três tipos de componentes no sistema de equações (4.1). O primeiro é o conjunto de coeficientes a_{ij} ; o segundo é o conjunto de variáveis $x_1, \dots, x_n$; e o último é o conjunto de termos constantes $d_1, \dots, d_m$. Se dispusermos os três conjuntos em três arranjos retangulares e os denominarmos, respectivamente, A , x e d (sem índices), então temos

Como um exemplo simples, dado o sistema de equações lineares

$$\begin{aligned} 6x_1 + 3x_2 - x_3 &= 22 \\ x_1 + 4x_2 - 2x_3 &= 12 \\ 4x_1 - x_2 + 5x_3 &= 10 \end{aligned} \tag{4.3}$$

podemos escrever

$$A = \begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad d = \begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} \quad (4.4)$$

Cada um dos três arranjos em (4.2) ou (4.4) constitui uma *matriz*.

Uma matriz é definida como um arranjo retangular de números, parâmetros ou variáveis. Os membros do arranjo, denominados *elementos* da matriz, usualmente estão entre colchetes, como em (4.2), ou, às vezes, entre parênteses ou entre linhas verticais duplas: $\| \|$. Note que na matriz A (a *matriz de coeficientes* do sistema de equações) os elementos não estão separados por vírgulas, mas apenas por espaços em branco. Usando uma notação abreviada, o conjunto na matriz A pode ser escrito de modo mais simples como

$$A = [a_{ij}] \quad \begin{pmatrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{pmatrix}$$

Considerando que a localização de cada elemento em uma matriz está inequivocamente fixada pelo índice, toda matriz é um conjunto ordenado.

Vetores como matrizes especiais

O número de linhas e o número de colunas em uma matriz, em conjunto, definem a *dimensão* da matriz. Visto que a matriz A em (4.2) contém m linhas e n colunas, diz-se que sua dimensão é $m \times n$ (lê-se “ m por n ”). É importante lembrar que o número de linhas vem sempre antes do número de colunas, de acordo com o modo como estão ordenados os dois índices em a_{ij} . No caso especial em que $m = n$, a matriz é denominada uma *matriz quadrada*. Assim, a matriz A em (4.4) é uma matriz quadrada 3×3 .

Algumas matrizes podem conter somente uma coluna, como x e d em (4.2) ou (4.4). Essas matrizes recebem o nome especial de *vetores coluna*. Em (4.2), a dimensão de x é $n \times 1$, e a de d é $m \times 1$; em (4.4), ambas, x e d , são 3×1 . Se disposto as variáveis x_j em um arranjo horizontal, no entanto, resultaria uma matriz $1 \times n$, que é denominada um *vetor linha*. Para a finalidade de notação, um vetor linha muitas vezes é distinguido de um vetor coluna pela utilização de um símbolo, uma aspa simples, denominado “linha”:

$$x' = [x_1 \quad x_2 \quad \dots \quad x_n]$$

O leitor pode observar que um vetor (seja vetor linha ou vetor coluna) é meramente um conjunto ordenado de n elementos e, como tal, às vezes pode ser interpretado como um ponto em um espaço n -dimensional. Por sua vez, a matriz A $m \times n$ pode ser interpretada como um conjunto ordenado de m vetores linhas ou como um conjunto ordenado de n vetores colunas. Essas idéias serão retomadas no Capítulo 5.

Uma questão de interesse mais imediato é como a notação matricial pode nos habilitar, como prometido, a expressar um sistema de equações de um modo compacto. Com as matrizes definidas em (4.4), podemos expressar o sistema de equações (4.3) simplesmente como

$$Ax = d$$

De fato, se atribuirmos a A , x e d os significados em (4.2), então até mesmo o sistema geral de equações em (4.1) pode ser escrito como $Ax = d$. Assim, a compacidade dessa notação é evidente.

Contudo, a equação $Ax = d$ sugere, no mínimo, duas perguntas. Como multiplicamos duas matrizes A e x ? O que quer dizer a igualdade de Ax e d ? Visto que matrizes envolvem blocos inteiros de números, as operações algébricas familiares definidas para números isolados não são aplicáveis diretamente e há necessidade de um novo conjunto de regras operacionais.

EXERCÍCIO 4.1

1. Escreva o modelo de mercado (3.1) no formato (4.1) e demonstre que, se as três variáveis forem dispostas na ordem Q_d , Q_s e P , a matriz de coeficientes será

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & b \\ 1 & 1 & -d \end{bmatrix}$$

Como você escreveria o vetor de constantes?

2. Escreva o modelo de mercado (3.12) no formato (4.1) com as variáveis colocadas na seguinte ordem: Q_{d1} , Q_{s1} , Q_{d2} , Q_{s2} , P_1 , P_2 . Escreva a matriz de coeficientes, o vetor de variáveis e o vetor de constantes.
3. O modelo de mercado (3.6) pode ser escrito na forma (4.1)? Por quê?
4. Escreva o modelo de renda nacional (3.23) na forma (4.1), tendo Y como a primeira variável. Escreva a matriz de coeficientes e o vetor de constantes.
5. Escreva o modelo de renda nacional do Exercício 3.5–1 na forma (4.1), com as variáveis na ordem Y , T e C . [Sugestão: Cuidado com a expressão multiplicativa $b(Y - T)$ na função de consumo].

4.2 Operações com matrizes

Antes de mais nada, vamos definir a palavra *igualdade*. Diz-se que duas matrizes, $A = [a_{ij}]$ e $B = [b_{ij}]$, são iguais se e somente se elas tiverem a mesma dimensão e elementos idênticos nas localizações correspondentes no arranjo. Em outras palavras, $A = B$ se e somente se $a_{ij} = b_{ij}$ para todos os valores de i e j . Assim, achamos, por exemplo,

$$\begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 2 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$$

Em outro exemplo, se $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$, então $x = 7$ e $y = 4$.

Adição e subtração de matrizes

Duas matrizes podem ser somadas se e somente se tiverem a mesma dimensão. Quando esse requisito dimensional for satisfeito, diz-se que as matrizes são *conformes para adição*. Nesse caso, a adição de $A = [a_{ij}]$ e $B = [b_{ij}]$ é definida como a adição de cada par de elementos correspondentes.

EXEMPLO 1

$$\begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 2 & 8 \end{bmatrix}$$

EXEMPLO 2

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \end{bmatrix}$$

Em geral, podemos enunciar a regra da seguinte maneira:

$$[a_{ij}] + [b_{ij}] = [c_{ij}] \quad \text{onde } c_{ij} = a_{ij} + b_{ij}$$

Note que a matriz soma $[c_{ij}]$ deve ter a mesma dimensão das matrizes componentes $[a_{ij}]$ e $[b_{ij}]$.

A operação de subtração $A - B$ pode ser definida, de maneira semelhante, se e somente se A e B tiverem a mesma dimensão. A operação implica o resultado

$$[a_{ij}] - [b_{ij}] = [d_{ij}] \quad \text{onde } d_{ij} = a_{ij} - b_{ij}$$

EXEMPLO 3

$$\begin{bmatrix} 19 & 3 \\ 2 & 0 \end{bmatrix} - \begin{bmatrix} 6 & 8 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 19-6 & 3-8 \\ 2-1 & 0-3 \end{bmatrix} = \begin{bmatrix} 13 & -5 \\ 1 & -3 \end{bmatrix}$$

A operação de subtração $A - B$ pode ser considerada alternativamente como uma operação de adição envolvendo a matriz A e uma outra matriz $(-1)B$. Contudo, isso suscita a questão do que significa a multiplicação de uma matriz por um número isolado (aqui, -1).

Multiplicação escalar

Multiplicar uma matriz por um número – ou, na terminologia de álgebra matricial, por um *escalar* – é multiplicar *cada* elemento da matriz em questão pelo dado escalar.

EXEMPLO 4

$$7 \begin{bmatrix} 3 & -1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 21 & -7 \\ 0 & 35 \end{bmatrix}$$

EXEMPLO 5

$$\frac{1}{2} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{21} & \frac{1}{2}a_{22} \end{bmatrix}$$

Com esses exemplos, o princípio racional do nome escalar deve ficar claro, pois o escalar faz o valor da matriz aumentar (“escalar” para cima) ou diminuir (“escalar” para baixo) por um certo múltiplo. O escalar também pode, é claro, ser um número negativo.

EXEMPLO 6

$$-1 \begin{bmatrix} a_{11} & a_{12} & d_1 \\ a_{21} & a_{22} & d_2 \end{bmatrix} = \begin{bmatrix} -a_{11} & -a_{12} & -d_1 \\ -a_{21} & -a_{22} & -d_2 \end{bmatrix}$$

Note que, se a matriz à esquerda representar os coeficientes e os termos constantes nas equações simultâneas

$$a_{11}x_1 + a_{12}x_2 = d_1$$

$$a_{21}x_1 + a_{22}x_2 = d_2$$

então a multiplicação pelo escalar -1 será o mesmo que multiplicar ambos os lados da equação por -1 , trocando, assim, o sinal de todos os termos no sistema.

Multiplicação de matrizes

Enquanto um escalar pode ser usado para multiplicar uma matriz de qualquer dimensão, a multiplicação de duas matrizes depende da satisfação de um requisito dimensional diferente.

Suponha que, dadas duas matrizes, A e B , queiramos achar o produto AB . A condição de conformidade para a multiplicação é que a dimensão da *coluna* de A (a matriz “guia” na expressão AB) deve ser igual à dimensão da linha de B (a matriz “guiada”).

Por exemplo, se

$$\underset{(1 \times 2)}{A} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \quad \underset{(2 \times 3)}{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \quad (4.5)$$

então o produto AB é definido, já que A tem *duas colunas* e B tem *duas linhas* – exatamente o mesmo número.[†] Isso pode ser verificado imediatamente comparando o *segundo* número do indicador de dimensão de A , que é (1×2) , com o *primeiro* número do indicador de dimensão de B , (2×3) . Por outro lado, o produto inverso BA não é definido nesse caso, porque B (agora a matriz guia) tem *três* colunas, enquanto A (a matriz guiada) tem somente *uma* linha; por conseguinte, a condição de conformidade foi violada.

Em geral, se a dimensão de A for $m \times n$ e a dimensão de B for $p \times q$, o produto das matrizes, AB , será definido se e somente se $n = p$. Além do mais, se definido, a matriz produto AB terá a dimensão $m \times q$ – o mesmo número de *linhas* da matriz guia A e o mesmo número de *colunas* da matriz guiada B . Para as matrizes dadas em (4.5), AB será 1×3 .

Resta definir o procedimento exato de multiplicação. Para essa finalidade, vamos utilizar como ilustração as matrizes A e B em (4.5). Visto que o produto AB é definido e sua dimensão esperada é 1×3 , podemos escrever, de modo geral (usando o símbolo C em vez de c' para o vetor linha), que

$$AB = C = [c_{11} \quad c_{12} \quad c_{13}]$$

Cada elemento na matriz produto C , denotado por c_{ij} , é definido como uma soma de produtos a ser calculada a partir dos elementos na i -ésima *linha* da matriz guia A e dos elementos na j -ésima *coluna* na matriz guiada B . Para achar c_{11} , por exemplo, devemos tomar a *primeira linha* de A (já que $i = 1$) e a *primeira coluna* de B (já que $j = 1$) – como mostra a parte superior da Figura 4.1 – e então reunir os elementos em pares em seqüência, multiplicar cada par e tomar a soma dos produtos resultantes, para obter

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} \quad (4.6)$$

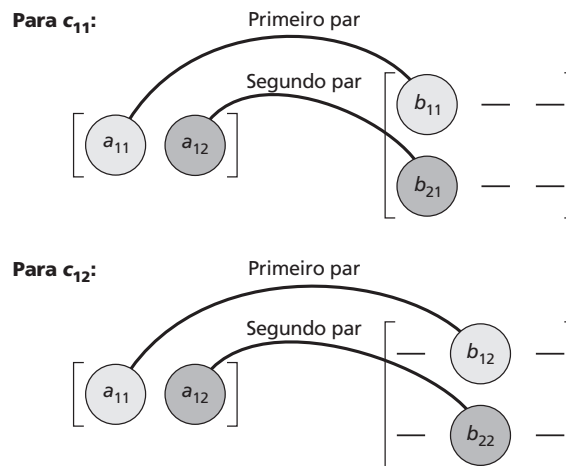
De forma semelhante, para c_{12} , pegamos a *primeira linha* de A (já que $i = 1$) e a *segunda coluna* de B (já que $j = 2$) e calculamos a soma indicada de produtos – de acordo com o painel inferior da Figura 4.1 – como se segue:

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} \quad (4.6')$$

Pelo mesmo critério, devemos ter também

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} \quad (4.6'')$$

FIGURA 4.1



[†] A matriz A , por ser um vetor linha, normalmente seria denotada por a' . Aqui usamos o símbolo A para salientar o fato de que a regra de multiplicação que está sendo explicada se aplica a matrizes em geral e não somente ao produto de um vetor e uma matriz.

É o requisito específico de formação de pares nesse processo que exige a compatibilização entre a dimensão da coluna da matriz guia e a dimensão da linha da matriz guiada antes de executar a multiplicação.

O procedimento de multiplicação ilustrado na Figura 4.1 também pode ser descrito utilizando o conceito de *produto interno* de dois vetores. Dados dois vetores u e v com n elementos cada, digamos, $(u_1, u_2, \dots, u_n)$ e $(v_1, v_2, \dots, v_n)$, dispostos *ou* como duas linhas *ou* como duas colunas *ou* como uma linha e uma coluna, seu produto interno, denotado como $u \cdot v$ (com um ponto no meio), é definido como

$$u \cdot v = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

Essa é a soma dos produtos de elementos correspondentes e, por conseguinte, o produto interno dos dois vetores é um escalar.

EXEMPLO 7

Se, após uma rodada de compras, dispusermos as quantidades compradas de n bens como um vetor linha $Q' = [Q_1 \ Q_2 \ \dots \ Q_n]$ e listarmos os preços desses bens em um vetor de preços $P' = [P_1 \ P_2 \ \dots \ P_n]$, então o produto interno desses dois vetores é

$$Q' \cdot P' = Q_1 P_1 + Q_2 P_2 + \dots + Q_n P_n = \text{custo total da compra}$$

Usando esse conceito, podemos descrever o elemento c_{ij} na matriz produto $C = AB$ simplesmente como o produto interno da i -ésima linha da matriz guia A e da j -ésima coluna da matriz guiada B . Examinando a Figura 4.1, podemos facilmente verificar a validade dessa descrição.

A regra de multiplicação que acabamos de esboçar se aplica com igual validade quando as dimensões de A e B são outras que não as ilustradas na Figura 4.1; o único pré-requisito é que seja cumprida a condição de conformidade.

EXEMPLO 8

Dados

$$A_{(3 \times 2)} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \\ 4 & 0 \end{bmatrix} \quad \text{e} \quad B_{(2 \times 1)} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

calcule AB . O produto AB é, de fato, definido, porque A tem duas colunas e B tem duas linhas. Sua matriz produto deve ser 3×1 , um vetor coluna:

$$AB = \begin{bmatrix} 1(5) + 3(9) \\ 2(5) + 8(9) \\ 4(5) + 0(9) \end{bmatrix} = \begin{bmatrix} 32 \\ 82 \\ 20 \end{bmatrix}$$

EXEMPLO 9

Dados

$$A_{(3 \times 3)} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix} \quad \text{e} \quad B_{(3 \times 3)} = \begin{bmatrix} 0 & -\frac{1}{5} & \frac{3}{10} \\ -1 & \frac{1}{5} & \frac{7}{10} \\ 0 & \frac{2}{5} & -\frac{1}{10} \end{bmatrix}$$

calcule AB . A mesma regra de multiplicação agora resulta em uma matriz produto muito especial:

$$AB = \begin{bmatrix} 0+1+0 & -\frac{3}{5}-\frac{1}{5}+\frac{4}{5} & \frac{9}{10}-\frac{7}{10}-\frac{2}{10} \\ 0+0+0 & -\frac{1}{5}+0+\frac{6}{5} & \frac{3}{10}+0-\frac{3}{10} \\ 0+0+0 & -\frac{4}{5}+0+\frac{4}{5} & \frac{12}{10}+0-\frac{2}{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Essa última matriz – uma matriz quadrada com números 1 em sua *diagonal principal* (a diagonal que vai da esquerda para a direita (noroeste para sudeste), e números zero no restante – exemplifica o tipo importante de matriz denominado *matriz identidade*. Esse assunto será mais detalhado na Seção 4.5.

EXEMPLO 10 Agora vamos considerar a matriz A e o vetor x como definidos em (4.4) e calcular Ax . A matriz produto é um vetor coluna 3×1 :

$$Ax = \begin{bmatrix} 6 & 3 & 1 \\ 1 & 4 & -2 \\ 4 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6x_1 + 3x_2 + x_3 \\ x_1 + 4x_2 - 2x_3 \\ 4x_1 - x_2 + 5x_3 \end{bmatrix}$$

(3×3) (3×1) (3×1)

Nota: O produto à direita é um vetor *coluna*, não obstante sua aparência corpulenta! Quando escrevemos $Ax = d$, portanto, temos

$$\begin{bmatrix} 6x_1 + 3x_2 + x_3 \\ x_1 + 4x_2 - 2x_3 \\ 4x_1 - x_2 + 5x_3 \end{bmatrix} = \begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix}$$

que, segundo a definição de igualdade de matrizes, é equivalente ao sistema de equações em (4.3).

Observe que, para usar a notação matricial $Ax = d$, é preciso, por causa da condição de conformidade, dispor as variáveis x_j em um vetor *coluna*, mesmo que essas variáveis estejam listadas em ordem horizontal no sistema de equações original.

EXEMPLO 11 O modelo mais simples de renda nacional de duas variáveis endógenas Y e C ,

$$Y = C + I_0 + G_0$$

$$C = a + bY$$

pode ser rearranjado no formato padrão (4.1) como se segue:

$$Y - C = I_0 + G_0$$

$$-bY + C = a$$

Daí, a matriz de coeficientes A , o vetor de variáveis x e o vetor de constantes d são

$$A_{(2 \times 2)} = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix} \quad x_{(2 \times 1)} = \begin{bmatrix} Y \\ C \end{bmatrix} \quad d_{(2 \times 1)} = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

Vamos verificar se esse sistema pode ser expresso pela equação $Ax = d$.

Pela regra de multiplicação de matrizes, temos

$$Ax = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix} \begin{bmatrix} Y \\ C \end{bmatrix} = \begin{bmatrix} 1(Y) + (-1)(C) \\ -b(Y) + 1(C) \end{bmatrix} = \begin{bmatrix} Y - C \\ -bY + C \end{bmatrix}$$

Assim, a equação matricial $Ax = d$ nos dá

$$\begin{bmatrix} Y - C \\ -bY + C \end{bmatrix} = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

Visto que igualdade de matrizes significa a igualdade entre elementos correspondentes, é claro que a equação $Ax = d$ representa precisamente o sistema de equações original expresso na forma (4.1).

A questão da divisão

Ao passo que matrizes, assim como números, podem ser somadas, subtraídas e multiplicadas – sujeitas às condições de conformidade –, não é possível dividir uma matriz por outra. Isto é, não podemos escrever A/B .

Para dois números a e b , o quociente a/b (com $b \neq 0$) pode ser escrito, alternativamente, como ab^{-1} ou $b^{-1}a$, onde b^{-1} representa o *inverso* ou *recíproco* de b . Como $ab^{-1} = b^{-1}a$, a expressão de quociente a/b pode ser usada para representar ambos, ab^{-1} e $b^{-1}a$. O caso de matrizes é diferente. Aplicando o conceito de inversos a matrizes podemos, em certos casos (ver Seção 4.6), definir uma matriz B^{-1} que é o inverso da matriz B . Mas, da discussão da condição de conformidade, conclui-se que, se AB^{-1} for definido, não pode haver nenhuma segurança de que $B^{-1}A$ também seja definido. Mesmo que AB^{-1} e $B^{-1}A$ sejam, de fato, ambos definidos, ainda assim poderão não representar o mesmo produto. Por conseguinte, a expressão A/B não pode ser usada sem ambigüidade e deve ser evitada. Em vez disso, você deve especificar se está se referindo a AB^{-1} ou a $B^{-1}A$ – contanto que o inverso B^{-1} realmente exista e que o produto de matrizes em questão seja definido. Matrizes inversas serão discutidas com mais detalhes na Seção 4.6.

A notação Σ

A utilização de símbolos com índices não somente ajuda a designar as localizações de parâmetros e variáveis, mas também se presta a uma notação abreviada flexível para denotar soma de termos, como as que surgem durante o processo de multiplicação de matrizes.

A notação abreviada para somatório utiliza a letra grega Σ (sigma, de “soma”). Por exemplo, para expressar a soma de x_1 , x_2 e x_3 , podemos escrever

$$x_1 + x_2 + x_3 = \sum_{j=1}^3 x_j$$

que se lê “a soma de x_j quando j vai de 1 a 3”. O símbolo j , denominado *índice do somatório* assume somente valores inteiros. A expressão x_j representa o *somando* (aquele que deve ser somado) e é, com efeito, uma função de j . Além da letra j , índices de somatórios também são comumente notados por i ou k , tal como em

$$\sum_{i=3}^7 x_i = x_3 + x_4 + x_5 + x_6 + x_7$$

$$\sum_{k=0}^n x_k = x_0 + x_1 + \cdots + x_n$$

A aplicação da notação Σ pode ser prontamente estendida a casos em que o termo x é precedido de um coeficiente ou em que cada termo da soma é elevado a uma potência inteira positiva. Por exemplo, podemos escrever:

$$\sum_{j=1}^3 ax_j = ax_1 + ax_2 + ax_3 = a(x_1 + x_2 + x_3) = a \sum_{j=1}^3 x_j$$

$$\sum_{j=1}^3 a_j x_j = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\sum_{i=0}^n a_i x^i = a_0 x^0 + a_1 x^1 + a_2 x^2 + \cdots + a_n x^n$$

$$= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

O último exemplo, em particular, mostra que a expressão $\sum_{i=0}^n a_i x^i$ pode, de fato, ser usada como notação abreviada para a função polinomial geral de (2.4).

Aproveitamos para mencionar que, sempre que o contexto da discussão não der margem a nenhuma ambigüidade quanto à faixa do somatório, o símbolo $\sum$ pode ser usado sozinho, sem nenhum índice (como $\sum x_i$), ou com apenas o índice inferior (como $\sum x_i$).

Vamos aplicar a notação abreviada $\sum$ à multiplicação de matrizes. Em (4.6), (4.6') e (4.6''), cada elemento da matriz produto $C = AB$ é definido como uma soma de termos, que agora podem ser escritos da seguinte maneira:

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} = \sum_{k=1}^2 a_{1k}b_{k1}$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} = \sum_{k=1}^2 a_{1k}b_{k2}$$

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} = \sum_{k=1}^2 a_{1k}b_{k3}$$

Em cada caso, o primeiro índice de c_{1j} é refletido no primeiro índice de a_{1k} , e o segundo índice de c_{1j} é refletido no segundo índice de b_{kj} na expressão $\sum$. O índice k , por outro lado, é um índice “fictício”; serve para indicar qual par de elementos específico está sendo multiplicado, mas não aparece no símbolo c_{1j} .

Estendendo isso para a multiplicação de uma matriz $m \times n$, $A = [a_{ik}]$, e para uma matriz $n \times p$, $B = [b_{kj}]$, agora podemos escrever os elementos da matriz produto $m \times p$, $AB = C = [c_{ij}]$, como

$$c_{11} = \sum_{k=1}^n a_{1k}b_{k1} \quad c_{12} = \sum_{k=1}^n a_{1k}b_{k2} \quad \dots$$

ou, de maneira mais geral,

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} \quad \left(\begin{array}{l} i = 1, 2, \dots, m \\ j = 1, 2, \dots, p \end{array} \right)$$

Essa última equação representa um outro modo de enunciar a regra de multiplicação de matrizes definida anteriormente.

EXERCÍCIO 4.2

1. Dados $A = \begin{bmatrix} 7 & -1 \\ 6 & 9 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 4 \\ 3 & -2 \end{bmatrix}$ e $C = \begin{bmatrix} 8 & 3 \\ 6 & 1 \end{bmatrix}$, calcule:

- (a) $A + B$ (b) $C - A$ (c) $3A$ (d) $4B + 2C$

2. Dados $A = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 \\ 3 & 8 \end{bmatrix}$ e $C = \begin{bmatrix} 7 & 2 \\ 6 & 3 \end{bmatrix}$:

- (a) AB é definido? Calcule AB . É possível calcular BA ? Por quê?
 (b) BC é definido? Calcule BC . CB é definido? Se for, calcule CB . É verdade que $BC = CB$?
 3. Tendo como base as matrizes dadas no Exemplo 9, o produto BA é definido? Se for, calcule o produto. Nesse caso, temos $AB = BA$?
 4. Calcule as matrizes produtos nos seguintes casos (em cada caso, acrescente um indicador de dimensão embaixo de cada matriz):

$$(a) \begin{bmatrix} 0 & 2 & 0 \\ 3 & 0 & 4 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 1 \\ 3 & 5 \end{bmatrix}$$

$$(c) \begin{bmatrix} 3 & 5 & 0 \\ 4 & 2 & -7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$(b) \begin{bmatrix} 6 & 5 & -1 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 5 & 2 \\ 0 & 1 \end{bmatrix}$$

$$(d) \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} 7 & 0 \\ 0 & 2 \\ 1 & 4 \end{bmatrix}$$

5. No Exemplo 7, se dispusermos as quantidades e preços como vetores colunas em vez de vetores linhas, $Q \cdot P$ é definido? Podemos expressar o custo total de compra como $Q \cdot P$? Como $Q' \cdot P$? Como $Q \cdot P'$?

6. Expanda as seguintes expressões de somatório:

$$(a) \sum_{i=2}^5 x_i$$

$$(d) \sum_{i=1}^n a_i x^{i-1}$$

$$(b) \sum_{i=5}^8 a_i x_i$$

$$(e) \sum_{i=0}^3 (x + i)^2$$

$$(c) \sum_{i=1}^4 b x_i$$

7. Escreva as seguintes expressões em notação Σ :

$$(a) x_1(x_1 - 1) + 2x_2(x_2 - 1) + 3x_3(x_3 - 1)$$

$$(b) a_2(x_3 + 2) + a_3(x_4 + 3) + a_4(x_5 + 4)$$

$$(c) \frac{1}{x} + \frac{1}{x^2} + \cdots + \frac{1}{x^n} \quad (x \neq 0)$$

$$(d) 1 + \frac{1}{x} + \frac{1}{x^2} + \cdots + \frac{1}{x^n} \quad (x \neq 0)$$

8. Mostre que as seguintes igualdades são verdadeiras:

$$(a) \left(\sum_{i=0}^n x_i \right) + x_{n+1} = \sum_{i=0}^{n+1} x_i$$

$$(b) \sum_{j=1}^n a b_j y_j = a \sum_{j=1}^n b_j y_j$$

$$(c) \sum_{j=1}^n (x_j + y_j) = \sum_{j=1}^n x_j + \sum_{j=1}^n y_j$$

4.3 Notas sobre operações vetoriais

Nas Seções 4.1 e 4.2, os vetores são considerados como um tipo especial de matriz. Como tal, eles se qualificam para a aplicação de todas as operações algébricas discutidas. Entretanto, devido a suas peculiaridades dimensionais, é interessante fazer alguns comentários adicionais sobre operações vetoriais.

Multiplicação de vetores

Um vetor coluna $m \times 1$, u , e um vetor linha $1 \times n$, v' , resultam em uma matriz produto uv' de dimensão $m \times n$.

EXEMPLO 1 Dados $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ e $v' = [1 \ 4 \ 5]$, podemos obter

$$uv' = \begin{bmatrix} 3(1) & 3(4) & 3(5) \\ 2(1) & 2(4) & 2(5) \end{bmatrix} = \begin{bmatrix} 3 & 12 & 15 \\ 2 & 8 & 10 \end{bmatrix}$$

Visto que cada linha em u consiste em somente um elemento, assim como cada coluna em v' , cada elemento de uv' revela-se um produto único, em vez de uma soma de produtos. O produto uv' é uma matriz 2×3 , mesmo que tenhamos começado com um par de vetores.

Por outro lado, dados um vetor linha $1 \times n$, u' , e um vetor coluna $n \times 1$, v , o produto $u'v$ terá dimensão 1×1 .

EXEMPLO 2 Dados $u' = [3 \ 4]$ e $v = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$, temos

$$u'v = [3(9) + 4(7)] = [55]$$

Como escrito, $u'v$ é uma matriz a despeito do fato de apenas um elemento estar presente. Entretanto, matrizes 1×1 se comportam exatamente como escalares no que diz respeito à adição e à multiplicação: $[4] + [8] = [12]$, exatamente como $4 + 8 = 12$; e $[3][7] = [21]$, exatamente como $3(7) = 21$. Além disso, matrizes 1×1 não possuem nenhuma propriedade importante que os escalares não tenham. De fato, há uma correspondência um a um entre o conjunto de todos os escalares e o conjunto de todas as matrizes 1×1 cujos elementos são escalares. Por essa razão, podemos redefinir $u'v$ como o *escalar* correspondente à matriz produto 1×1 . Conseqüentemente, no Exemplo 2, podemos escrever $u'v = 55$. Esse produto é denominado um *produto escalar*.[†] Lembre-se, contudo, que, enquanto uma matriz 1×1 pode ser tratada como um escalar, um escalar não pode ser substituído à vontade por uma matriz 1×1 , caso seja preciso continuar o cálculo, porque podem surgir complicações em relação às condições de conformidade.

EXEMPLO 3 Dado um vetor linha $u' = [3 \ 6 \ 9]$, ache $u'u$. Uma vez que u é meramente o vetor coluna com os elementos de u' arranjados verticalmente, temos

$$u'u = [3 \ 6 \ 9] \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = (3)^2 + (6)^2 + (9)^2$$

onde omitimos os colchetes da matriz produto 1×1 à direita. Note que o produto $u'u$ dá a soma dos quadrados dos elementos de u .

Em geral, se $u' = [u_1 \ u_2 \ \dots \ u_n]$, então $u'u$ será a soma dos quadrados (um escalar) dos elementos u_j :

$$u'u = u_1^2 + u_2^2 + \dots + u_n^2 = \sum_{j=1}^n u_j^2$$

Se tivéssemos calculado o produto interno $u \cdot u$ (ou $u' \cdot u'$), é claro que teríamos obtido exatamente o mesmo resultado.

Concluindo, é importante distinguir os significados de uv' (uma matriz maior que 1×1) e $u'v$ (uma matriz 1×1 , ou um escalar). Observe, em particular, que um produto escalar deve ter um vetor *linha* como matriz guia e um vetor *coluna* como matriz guiada; caso contrário, o produto não pode ser 1×1 .

Interpretação geométrica de operações vetoriais

Como mencionado anteriormente, um vetor linha ou um vetor coluna com n elementos (daqui em diante denominado um *vetor* n) pode ser visto como uma n -upla e, conseqüentemente, como um

[†] O conceito de produto escalar é, portanto, análogo ao conceito de produto interno de dois vetores, cada qual com o mesmo número de elementos, que também resulta em um escalar. Lembre-se, entretanto, que o produto interno está isento da condição de conformidade para a multiplicação, de modo que podemos escrevê-lo como $u \cdot v$. No caso de um produto escalar (cuja notação não tem um ponto entre os dois símbolos de vetor), por outro lado, podemos expressá-lo apenas como um vetor linha multiplicado por um vetor coluna, tendo o vetor linha como guia.

ponto em um espaço de n dimensões (daqui em diante denominado um *espaço n*). Vamos detalhar mais essa idéia. Na Figura 4.2a, um colocado $(3, 2)$ é colocado em um espaço 2 e é denominado u .

Isso é a contraparte geométrica do vetor $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ ou do vetor $u' = [3 \ 2]$, ambos indicando, nesse

contexto, um único e mesmo par ordenado. Se desenharmos uma seta (um segmento de reta com direção definida) do ponto de origem $(0, 0)$ até o ponto u , ela especificará o único caminho reto para se chegar ao ponto de destino, u , partindo do ponto de origem. Uma vez que existe uma única seta para cada ponto, podemos considerar que o vetor u é representado graficamente *ou* pelo ponto $(3, 2)$ *ou* pela seta correspondente. Essa seta, que parte da origem $(0, 0)$ como o ponteiro de um relógio, tem um comprimento definido e uma direção definida, e é denominada um *vetor raio*.

Segundo essa nova interpretação de um vetor, torna-se possível atribuir significados geométricos: (a) à multiplicação escalar de um vetor; (b) à adição e à subtração de vetores; e, de modo mais geral, (c) à denominada combinação linear de vetores.

Primeiro, se colocarmos o vetor $\begin{bmatrix} 6 \\ 4 \end{bmatrix} = 2u$ na Figura 4.2a, a seta resultante ficará sobreposta à antiga, mas será duas vezes mais longa. De fato, a multiplicação do vetor u por qualquer escalar k produzirá uma seta sobreposta, mas a localização da ponta da seta será outra, a menos que $k = 1$. Se o multiplicador escalar for $k > 1$, a seta será aumentada (escalará para cima); se $0 < k < 1$, a seta será encurtada (escalará para baixo); se $k = 0$, a seta encolherá até o ponto de origem – que representa um *vetor nulo*, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Um multiplicador escalar negativo até mesmo inverterá a direção da seta.

Se o vetor u for multiplicado por -1 , por exemplo, obtemos $-u = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$, que é representado graficamente na Figura 4.2b como uma seta com o mesmo comprimento de u , mas com direção diametralmente oposta.

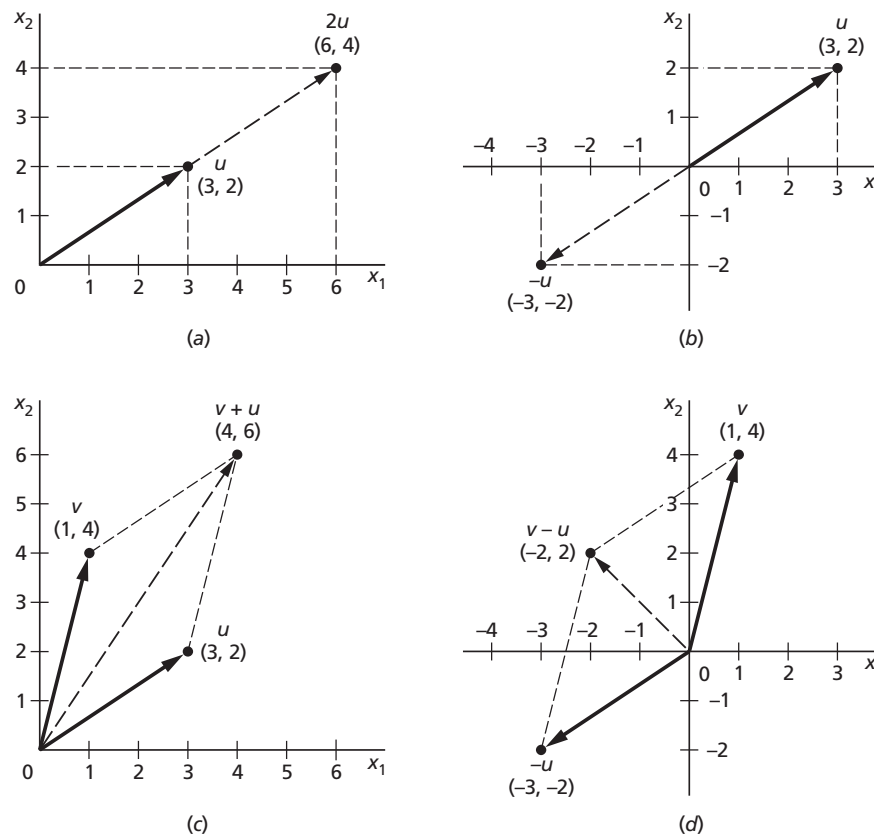


FIGURA 4.2

Em seguida, considere a adição de dois vetores, $v = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ e $u = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. A soma $v + u = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ pode ser

representada pela seta tracejada na Figura 4.2c. Contudo, se construirmos um paralelogramo no qual dois lados são os dois vetores u e v (setas em linhas cheias), a diagonal do paralelogramo será exatamente a seta que representa a soma vetorial $v + u$. Em geral, uma soma vetorial pode ser obtida geometricamente de um paralelogramo. Além disso, esse método também pode nos dar a *diferença vetorial* $v - u$, já que esta é equivalente à soma de v e $(-1)u$. Na Figura 4.2d, em primeiro lugar, reproduzimos o vetor v e o vetor negativo $-u$ dos diagramas c e b, respectivamente, e então construímos um paralelogramo. A diagonal resultante representa a diferença vetorial $v - u$.

Uma simples extensão desses resultados é suficiente para interpretar geometricamente uma combinação linear (isto é, uma soma ou diferença linear) de vetores. Considere o caso simples de

$$3v + 2u = 3 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 16 \end{bmatrix}$$

O aspecto da multiplicação escalar dessa operação envolve a realocização das respectivas pontas das setas dos dois vetores v e u , e o aspecto da adição pede a construção de um paralelogramo. Além dessas duas operações geométricas básicas, não há nada de novo em uma combinação linear de vetores. Isso é válido mesmo se houver mais termos na combinação linear, como em

$$\sum_{i=1}^n k_i v_i = k_1 v_1 + k_2 v_2 + \cdots + k_n v_n$$

onde k_i representa um conjunto de escalares, mas os símbolos v_i , com índices, agora denotam um conjunto de vetores. Para formar essa soma, os dois primeiros termos podem ser somados em primeiro lugar e, então, a soma resultante adicionada ao terceiro e assim por diante até que todos os termos sejam incluídos.

Dependência linear

Diz-se que um conjunto de vetores $v_1, \dots, v_n$ é *linearmente dependente* se (e somente se) qualquer um deles puder ser expresso como uma combinação linear dos vetores restantes; caso contrário, são *linearmente independentes*.

EXEMPLO 4 Os três vetores $v_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$ e $v_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ são linearmente dependentes porque v_3 é uma combinação linear de v_1 e v_2 :

$$3v_1 - 2v_2 = \begin{bmatrix} 6 \\ 21 \end{bmatrix} - \begin{bmatrix} 2 \\ 16 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} = v_3$$

Note que essa última equação pode ser expressa alternativamente como

$$3v_1 - 2v_2 - v_3 = 0$$

onde $0 \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ representa um vetor nulo (também denominado *vetor zero*).

EXEMPLO 5 Os dois vetores linhas $v'_1 = [5 \ 12]$ e $v'_2 = [10 \ 24]$ são linearmente dependentes porque

$$2v'_1 = 2 [5 \ 12] = [10 \ 24] = v'_2$$

O fato de um vetor ser um múltiplo de um outro vetor ilustra o caso mais simples de combinação linear. Observe, novamente, que essa última equação pode ser escrita equivalentemente como

$$2v_1' - v_2' = 0'$$

onde $0'$ representa o vetor linha nulo $[0 \ 0]$.

Com a introdução de vetores nulos, a dependência linear pode ser redefinida como segue. Um conjunto de vetores $m \ v_1, \dots, v_n$ é *linearmente dependente* se e somente se existir um conjunto de escalares $k_1, \dots, k_n$ (nem todos zeros) tal que

$$\sum_{i=1}^n k_i v_i = \underset{(m \times 1)}{0}$$

Por outro lado, se essa equação puder ser satisfeita *somente quando* $k_i = 0$ para todos os i , esses vetores são linearmente independentes.

O conceito de dependência linear admite também uma interpretação geométrica fácil. Dois vetores u e $2u$ – um múltiplo do outro – são obviamente dependentes. Geometricamente, na Figura 4.2a, suas setas estão sobre uma única reta. O mesmo é verdadeiro para os dois vetores dependentes u e $-u$ na Figura 4.2b. Ao contrário, os dois vetores u e v da Figura 4.2c são linearmente *independentes*, porque é impossível expressar um como múltiplo do outro. Geometricamente, suas setas não estão sobre uma única linha reta.

Quando são considerados mais de dois vetores no espaço 2, surge a seguinte conclusão significativa: uma vez encontrados dois vetores linearmente *independentes* no espaço 2 (digamos, u e v), todos os outros vetores naquele espaço poderão ser expressos como uma combinação linear desses dois (u e v). Nas Figuras 4.2c e d, já foi ilustrado como as duas combinações lineares simples $v + u$ e $v - u$ podem ser encontradas. Além do mais, aumentando, reduzindo e invertendo os dados vetores u e v e depois combinando esses vetores em vários paralelogramos, podemos gerar um número infinito de novos vetores, que esgotarão o conjunto de todos os vetores 2. Por causa disso, qualquer conjunto de três ou mais vetores 2 (três ou mais vetores em um espaço 2) tem de ser linearmente dependente. Dois deles podem ser independentes, mas o terceiro deve ser uma combinação linear dos dois primeiros.

Espaço vetorial

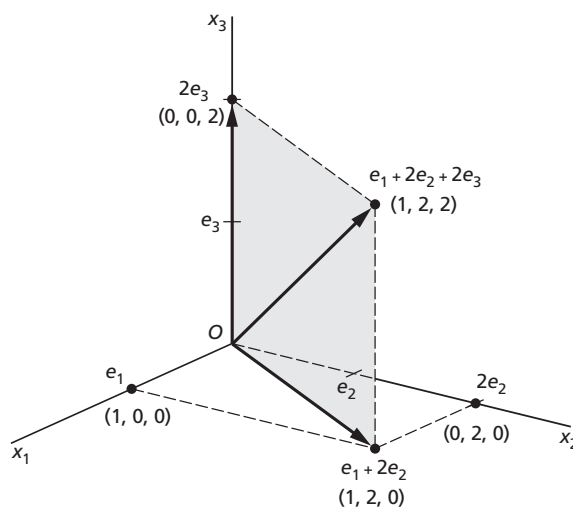
O total de vetores 2 gerados pelas várias combinações lineares de dois vetores independentes u e v constitui o *espaço vetorial* bidimensional. Uma vez que estamos tratando somente com vetores cujos elementos têm valores reais, esse espaço vetorial não é nenhum outro senão R^2 , o espaço 2 a que estivemos nos referindo o tempo todo. O espaço 2 não pode ser gerado por um único vetor 2 porque combinações lineares deste último não podem dar origem ao conjunto de vetores que fica sobre uma única reta. E a geração do espaço 2 tampouco requer mais do que dois vetores 2 linearmente independentes – de qualquer forma, seria impossível achar mais que dois.

Diz-se que os dois vetores linearmente independentes u e v *geram* o espaço 2. Diz-se, também, que eles constituem uma *base* para o espaço 2. Note que dissemos *uma* base e não *a* base, porque qualquer par de vetores 2 pode cumprir aquela finalidade contanto que sejam linearmente independentes. Em particular, considere os dois vetores $[1 \ 0]$ e $[0 \ 1]$, que são denominados *vetores unidade*. A representação gráfica do primeiro é uma seta ao longo do eixo horizontal e a do segundo é uma seta ao longo do eixo vertical. Como eles são linearmente independentes, podem servir como uma base para o espaço 2 e, na verdade, nós geralmente imaginamos o espaço 2 como o espaço abrangido por seus dois eixos, que nada mais são do que versões ampliadas dos dois vetores unidade.

Por analogia, o espaço vetorial tridimensional é a totalidade de vetores 3 e deve ser gerado por exatamente três vetores 3 linearmente independentes. Considere, como ilustração, o conjunto de três vetores unidade

$$e_1 \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 \equiv \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (4.7)$$

FIGURA 4.3



onde cada e_i é um vetor cujo i -ésimo elemento é 1 e todos os outros elementos são zeros.[†] Esses três vetores são obviamente linearmente independentes; de fato, suas setas ficam ao longo dos três eixos do espaço tridimensional na Figura 4.3. Assim, eles geram o espaço 3, o que implica que o espaço 3 inteiro (R^3 , em nossa estrutura) pode ser gerado por esses três vetores unidade.

Por exemplo, o vetor $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ pode ser considerado como a combinação linear $e_1 + 2e_2 + 2e_3$. Geometri-

camente, podemos primeiramente somar os vetores e_1 e $2e_2$ na Figura 4.3 pelo método do paralelogramo, de modo a obter o vetor representado pelo ponto $(1, 2, 0)$ no plano x_1x_2 , e então somar este último vetor a $2e_3$ – por meio do paralelogramo construído no plano vertical sombreado – para obter o resultado final desejado, no ponto $(1, 2, 2)$.

A extensão para espaço n deve ser óbvia. O espaço n pode ser definido como a totalidade de vetores n . Embora impossível de representar graficamente, ainda assim podemos imaginar o espaço n como o espaço abrangido por um total de n (elementos n) vetores unidade, todos linearmente independentes. Cada vetor n , sendo uma n -upla ordenada, representa um *ponto* no espaço n ou uma seta que se estende do ponto de origem (isto é, o vetor nulo de elementos n) até aquele ponto. E qualquer dado conjunto de n vetores n linearmente independentes é, de fato, capaz de gerar o espaço n inteiro. Visto que, em nossa discussão, cada elemento do vetor n está limitado a ser um número real, esse espaço n é, na verdade, R^n .

O espaço n a que nos referimos é, às vezes, denominado mais especificamente *espaço n euclidiano* (de Euclides). Para explicar esse último conceito, devemos, em primeiro lugar, comentar brevemente o conceito de *distância* entre dois pontos vetoriais. Para qualquer par de pontos vetoriais u e v em um dado espaço, a distância de u a v é alguma função de valor real

$$d = d(u, v)$$

com as seguintes propriedades: (1) quando u e v coincidem, a distância é zero; (2) quando os dois pontos são distintos, a distância de u a v e a distância de v a u são representadas por um número real idêntico; e (3) a distância entre u e v nunca é maior que a distância de u a w (um ponto distinto de u e v) mais a distância de w a v . Expresso simbolicamente,

$$d(u, v) = 0 \quad (\text{para } u = v)$$

$$d(u, v) = d(v, u) > 0 \quad (\text{para } u \neq v)$$

$$d(u, v) \leq d(u, w) + d(w, v) \quad (\text{para } w \neq u, v)$$

[†] O símbolo e pode ser associado com a palavra alemã *eins*, que significa “um”.

A última propriedade é conhecida como *desigualdade triangular* porque os três pontos u , v e w , juntos, usualmente definirão um triângulo.

Quando um espaço vetorial tem uma função de distância definida que cumpre as três propriedades citadas, é denominado um *espaço métrico*. Contudo, note que a distância $d(u, v)$ tem sido discutida apenas em termos gerais. Dependendo da forma específica atribuída à função d , pode resultar uma variedade de espaços métricos. O denominado espaço euclidiano é um tipo específico de espaço métrico com uma função de distância definida da seguinte maneira. Seja o ponto u a n -upla $(a_1, a_2, \dots, a_n)$ e o ponto v a n -upla $(b_1, b_2, \dots, b_n)$; então, a função da distância euclidiana é

$$d(u, v) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

onde admite-se que a raiz quadrada é positiva. Como pode ser facilmente verificado, essa função de distância específica satisfaz todas as três propriedades enumeradas anteriormente. Aplicada ao espaço bidimensional na Figura 4.2a, verifica-se que a distância entre os dois pontos $(6, 4)$ e $(3, 2)$ é

$$\sqrt{(6-3)^2 + (4-2)^2} = \sqrt{3^2 + 2^2} = \sqrt{13}$$

Esse resultado é considerado consistente com o *teorema de Pitágoras*, que diz que o comprimento da hipotenusa de um triângulo retângulo é igual à raiz quadrada (positiva) da soma dos quadrados dos comprimentos dos outros dois lados. Pois, se tomarmos $(6, 4)$ e $(3, 2)$ como u e v , e colocarmos um novo ponto w em $(6, 2)$, teremos, realmente, um triângulo retângulo cujos lados horizontal e vertical têm comprimentos iguais a 3 e 2, respectivamente, e cuja hipotenusa (a distância entre u e v) tem comprimento igual a $\sqrt{3^2 + 2^2} = \sqrt{13}$.

A função da distância euclidiana também pode ser expressa em termos da raiz quadrada de um produto escalar de dois vetores. Uma vez que u e v denotam as duas n -uplas $(a_1, \dots, a_n)$ e $(b_1, \dots, b_n)$, podemos escrever um vetor coluna $u - v$, com elementos $a_1 - b_1, a_2 - b_2, \dots, a_n - b_n$. O que vai embaixo do sinal de raiz quadrada na função da distância euclidiana é, é claro, simplesmente a soma dos quadrados desses n elementos que, em vista do Exemplo 3 desta seção, pode ser escrito como o produto escalar $(u - v)'(u - v)$. Por conseguinte, temos

$$d(u, v) = \sqrt{(u - v)'(u - v)}$$

EXERCÍCIO 4.3

- Dados $u' = [5 \ 1 \ 3]$, $v' = [3 \ 1 \ -1]$, $w' = [7 \ 5 \ 8]$ e $x' = [x_1 \ x_2 \ x_3]$, escreva os vetores colunas u , v , w e x , e calcule:

- | | | | |
|-----------|-----------|-----------|-----------|
| (a) uv' | (c) xx' | (e) $u'v$ | (g) $u'u$ |
| (b) uw' | (d) $v'u$ | (f) $w'x$ | (h) $x'x$ |

- Dados $w = \begin{bmatrix} 3 \\ 2 \\ 16 \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ e $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$:

- Quais dos seguintes são definidos: $w'x$, $x'y'$, xy' , $y'y$, zz' , yw' , $x \cdot y$?
- Calcule todos os produtos que são definidos.

- Tendo vendido n itens de mercadoria em quantidades $Q_1, \dots, Q_n$ e preços $P_1, \dots, P_n$, como você expressaria a receita total em (a) notação Σ ; e (b) notação vetorial?
- Dados dois vetores não-zero w_1 e w_2 , o ângulo θ ($0^\circ \leq \theta \leq 180^\circ$) que eles formam está relacionado ao produto escalar $w_1'w_2$ ($= w_2'w_1$) como segue:

$$\theta \text{ é um ângulo } \begin{cases} \text{agudo} \\ \text{reto} \\ \text{obtusos} \end{cases} \text{ se e somente se } w_1'w_2 \begin{cases} > \\ = \\ < \end{cases} 0$$

Verifique essa condição calculando o produto escalar de cada um dos seguintes pares de vetores (veja Figuras 4.2 e 4.3):

$$(a) w_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, w_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$(d) w_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, w_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

$$(b) w_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, w_2 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

$$(e) w_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, w_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

$$(c) w_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, w_2 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

5. Dados $u = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ e $v = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$, calcule as seguintes expressões graficamente:

$$(a) 2v$$

$$(c) u - v$$

$$(e) 2u + 3v$$

$$(b) u + v$$

$$(d) v - u$$

$$(f) 4u - 2v$$

6. Visto que o espaço tridimensional é abrangido pelos três vetores unidade definidos em (4.7), deve ser possível expressar qualquer outro vetor 3 como uma combinação linear de e_1 , e_2 e e_3 . Mostre que os seguintes vetores 3 podem ser expressos dessa forma:

$$(a) \begin{bmatrix} 4 \\ 7 \\ 0 \end{bmatrix}$$

$$(b) \begin{bmatrix} 25 \\ -2 \\ 1 \end{bmatrix}$$

$$(c) \begin{bmatrix} -1 \\ 6 \\ 9 \end{bmatrix}$$

$$(d) \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}$$

7. No espaço tridimensional euclidiano, qual é a distância entre os seguintes pontos?

$$(a) (3, 2, 8) \text{ e } (0, -1, 5)$$

$$(b) (9, 0, 4) \text{ e } (2, 0, -4)$$

8. A desigualdade triangular é escrita com o sinal de desigualdade *fraca* $\leq$, em vez do sinal de desigualdade estrita $<$. Sob quais circunstâncias se aplica a parte “=” da desigualdade?

9. Expresse o comprimento de um vetor raio, v , no espaço n euclidiano (isto é, a distância da origem até o ponto v) usando cada um dos seguintes:

$$(a) \text{ escalares}$$

$$(b) \text{ um produto escalar}$$

$$(c) \text{ um produto interno}$$

4.4 Leis comutativas, associativas e distributivas

Na álgebra escalar comum, as operações de adição e multiplicação obedecem às leis comutativa, associativa e distributiva como se segue:

Lei comutativa da adição:

$$a + b = b + a$$

Lei comutativa da multiplicação:

$$ab = ba$$

Lei associativa da adição:

$$(a + b) + c = a + (b + c)$$

Lei associativa da multiplicação:

$$(ab)c = a(bc)$$

Lei distributiva:

$$a(b + c) = ab + ac$$

Essas leis foram citadas durante a discussão das leis de nomes semelhantes aplicáveis à união e à interseção de conjuntos. A maioria dessas leis, mas não todas, também se aplicam a operações com matrizes – a exceção significativa é a lei comutativa da multiplicação.

Adição de matrizes

A adição de matrizes é comutativa, bem como associativa. Isso porque a adição de matrizes requer apenas a soma dos elementos correspondentes das duas matrizes e porque a ordem em que cada par de elementos é somado não tem importância. A propósito, nesse contexto, a operação

de subtração $A - B$ pode ser considerada simplesmente como a operação de adição $A + (-B)$ e, assim, não precisa ser discutida separadamente.

As leis comutativa e associativa podem ser enunciadas como se segue:

Lei comutativa

$$A + B = B + A$$

DEMONSTRAÇÃO $A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A$

EXEMPLO 1 Dados $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ e $B = \begin{bmatrix} 6 & 2 \\ 3 & 4 \end{bmatrix}$, verificamos que

$$A + B = B + A = \begin{bmatrix} 9 & 3 \\ 3 & 6 \end{bmatrix}$$

Lei associativa

$$(A + B) + C = A + (B + C)$$

DEMONSTRAÇÃO $(A + B) + C = [a_{ij} + b_{ij}] + [c_{ij}] = [a_{ij} + b_{ij} + c_{ij}]$
 $= [a_{ij}] + [b_{ij} + c_{ij}] = A + (B + C)$

EXEMPLO 2 Dados $v_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $v_2 = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$ e $v_3 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, verificamos que

$$(v_1 + v_2) - v_3 = \begin{bmatrix} 12 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

que é igual a

$$v_1 + (v_2 - v_3) = \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 7 \\ -4 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

Aplicada à combinação linear de vetores $k_1 v_1 + \dots + k_n v_n$, a lei associativa nos permite selecionar qualquer par de termos para adição (ou subtração) antes, em vez de ter de seguir a sequência na qual os n termos são listados.

Multiplicação de matrizes

A multiplicação de matrizes *não* é comutativa, isto é,

$$AB \neq BA$$

Como explicado anteriormente, mesmo quando AB é definido, BA pode não ser; mas, mesmo que ambos os produtos sejam definidos, a regra geral continua sendo $AB \neq BA$.

EXEMPLO 3 Seja $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ e $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$; então

$$AB = \begin{bmatrix} 1(0) + 2(6) & 1(-1) + 2(7) \\ 3(0) + 4(6) & 3(-1) + 4(7) \end{bmatrix} = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}$$

mas $BA = \begin{bmatrix} 0(1) - 1(3) & 0(2) - 1(4) \\ 6(1) + 7(3) & 6(2) + 7(7) \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ 27 & 40 \end{bmatrix}$

EXEMPLO 4 Seja u' 1×3 (um vetor linha); então, o vetor coluna correspondente, u , tem de ser 3×1 . O produto $u'u$ será 1×1 , mas o produto uu' será 3×3 . Assim, obviamente, $u'u \neq uu'$.

Em vista da regra geral $AB \neq BA$, os termos *pré-multiplicar* e *pós-multiplicar* são freqüentemente usados para especificar a ordem de multiplicação. No produto AB , diz-se que a matriz B é *pré-multiplicada* por A , e A é *pós-multiplicada* por B .

Contudo, existem exceções interessantes à regra $AB \neq BA$. Um desses casos é quando A é uma matriz quadrada e B é uma matriz identidade. Outro é quando A é o inverso de B , isto é, quando $A = B^{-1}$. Ambos os casos serão retomados mais adiante. Aqui é preciso observar, também, que a multiplicação escalar de uma matriz *obedece* à lei comutativa; assim, se k for um escalar, então

$$kA = Ak$$

Embora a multiplicação de matrizes não seja, em geral, comutativa, ela é associativa.

Lei associativa

$$(AB)C = A(BC) = ABC$$

Ao formar o produto ABC , a condição de conformidade deve ser naturalmente satisfeita por par *adjacente* de matrizes. Se A for $m \times n$ e se C for $p \times q$, então a conformidade requer que B seja $n \times p$.

$$\begin{matrix} A & B & C \\ (m \times n) & (n \times p) & (p \times q) \end{matrix}$$

Note que n e p aparecem duas vezes nos indicadores de dimensão. Se a condição de conformidade for cumprida, a lei associativa estabelece que qualquer par *adjacente* de matrizes pode ser multiplicado primeiro, contanto que o produto seja devidamente inserido no lugar exato do par original.

EXEMPLO 5 Se $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ e $A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$, então

$$x'Ax = x'(Ax) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & x_1 \\ 0 & a_{22} \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

Obtemos exatamente o mesmo resultado com

$$(x'A)x = \begin{bmatrix} a_{11}x_1 & a_{22}x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{22}x_2^2$$

No Exemplo 5, a matriz quadrada A tem os elementos não-nulos a_{11} e a_{22} na diagonal principal e zeros em todos os outros lugares. Essa matriz é denominada uma *matriz diagonal*. Quando uma matriz diagonal A aparece no produto $x'Ax$, o produto resultado dá uma soma de quadrados “ponderada” e os pesos para os termos x_1^2 e para os termos x_2^2 são fornecidos pelos elementos na diagonal de A . Esse resultado contrasta com o produto escalar $x'x$, que resulta em uma soma de quadrados simples (não ponderada).

EXEMPLO 6 Vamos definir o ideal econômico como o nível de receita nacional Y^0 ligado à taxa de inflação p^0 . E vamos supor que consideremos qualquer desvio positivo da receita real Y em relação a Y^0 tão igualmente indesejável quanto um desvio negativo da mesma grandeza, o mesmo acontecendo com o desvio entre a taxa de inflação real p e p^0 . Então, podemos escrever uma *função de perda social* tal como

$$\Lambda = \alpha(Y - Y^0)^2 + \beta(p - p^0)^2$$

onde α e β são os pesos atribuídos às duas fontes de perda social. Se os desvios em relação a Y forem considerados como o tipo de perda mais sério, então α deve ser maior que β . Note que elevar os desvios ao quadrado produz dois efeitos. Em primeiro lugar, elevando um desvio positivo ao quadrado ele receberá o mesmo valor de perda de um desvio negativo da mesma grandeza numérica. Em segundo lugar, elevar ao quadrado faz com que os desvios mais altos se destaquem de modo muito mais significativo que os desvios menores quando a perda social é medida. Tal função de perda social pode ser expressa, se desejarmos, pela matriz produto

$$\begin{bmatrix} Y - Y^0 & p - p^0 \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} Y - Y^0 \\ p - p^0 \end{bmatrix}$$

A multiplicação de matrizes também é distributiva.

Lei distributiva

$$A(B + C) = AB + AC \quad [\text{pré-multiplicação por } A]$$

$$(B + C)A = BA + CA \quad [\text{pós-multiplicação por } A]$$

Em cada caso, as condições de conformidade para a adição, bem como para a multiplicação, devem, é claro, ser observadas.

EXERCÍCIO 4.4

1. Dados $A = \begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 7 \\ 8 & 4 \end{bmatrix}$ e $C = \begin{bmatrix} 3 & 4 \\ 1 & 9 \end{bmatrix}$, verifique que

- (a) $(A + B) + C = A + (B + C)$
 (b) $(A + B) - C = A + (B - C)$

2. A subtração de uma matriz B pode ser considerada como a adição da matriz $(-1)B$. A lei comutativa da adição nos permite afirmar que $A - B = B - A$? Se a resposta for negativa, como você corrigiria essa afirmação?
3. Teste a lei associativa da multiplicação para as seguintes matrizes:

$$A = \begin{bmatrix} 5 & 3 \\ 0 & 5 \end{bmatrix} \quad B = \begin{bmatrix} -8 & 0 & 7 \\ 1 & 3 & 2 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 3 \\ 7 & 1 \end{bmatrix}$$

4. Demonstre que, para quaisquer dois escalares g e k :

- (a) $k(A + B) = kA + kB$
 (b) $(g + k)A = gA + kA$

(Nota: para *demonstrar* um resultado, você não pode usar exemplos específicos.)

5. Calcule $C = AB$ para (a), (b), (c) e (d).

(a) $A = \begin{bmatrix} 12 & 14 \\ 20 & 5 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 9 \\ 0 & 2 \end{bmatrix}$

(b) $A = \begin{bmatrix} 4 & 7 \\ 9 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 8 & 5 \\ 2 & 6 & 7 \end{bmatrix}$

(c) $A = \begin{bmatrix} 7 & 11 \\ 2 & 9 \\ 10 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 12 & 4 & 5 \\ 3 & 6 & 1 \end{bmatrix}$

(d) $A = \begin{bmatrix} 6 & 2 & 5 \\ 7 & 9 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 10 & 1 \\ 11 & 3 \\ 2 & 9 \end{bmatrix}$

- (e) Calcule (i) $C = AB$, e (ii) $D = BA$, para

$$A = \begin{bmatrix} -2 \\ 4 \\ 7 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 6 & -2 \end{bmatrix}$$

6. Demonstre que $(A + B)(C + D) = AC + AD + BC + BD$.
7. Se todos os quatro elementos da matriz A no Exemplo 5 fossem não-zero, $x'Ax$ ainda daria uma soma ponderada de quadrados? A lei associativa ainda se aplicaria?
8. Cite algumas situações ou contextos nos quais a noção de uma soma de quadrados ponderada ou não-ponderada pode ser relevante.

4.5 Matrizes identidade e matrizes nulas

Matrizes identidade

Já nos referimos anteriormente ao termo *matriz identidade*. Essa matriz é definida como uma matriz *quadrada* (repito: quadrada) que tem números 1 em sua diagonal principal e números 0 em todos os outros lugares. Ela é denotada pelo símbolo I , ou I_n , no qual o índice n serve para indicar a dimensão de sua linha, bem como de sua coluna. Assim,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Mas ambas também podem ser denotadas por I .

A importância desse tipo especial de matriz deve-se ao fato de que ela desempenha um papel semelhante ao do número 1 na álgebra escalar. Para qualquer número a , temos $1(a) = a(1) = a$. De maneira semelhante, temos, para qualquer matriz A ,

$$IA = AI = A \quad (4.8)$$

EXEMPLO 1 Seja $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix}$ então

$$IA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

$$AI = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \end{bmatrix} = A$$

Como A é 2×3 , a pré-multiplicação e a pós-multiplicação de A por I exigiria matrizes identidade de dimensões diferentes, a saber, I_2 e I_3 , respectivamente. Mas, no caso em que A é $n \times n$, então a mesma matriz identidade, I_n , pode ser usada, de modo que (4.8) torna-se $I_n A = A I_n$, ilustrando, assim, uma exceção à regra que diz que a multiplicação de matrizes não é comutativa.

A natureza especial das matrizes identidades possibilita *inserir* ou *remover* uma matriz identidade durante o processo de multiplicação sem afetar o produto de matrizes. Isso decorre diretamente de (4.8). Recordando a lei associativa, temos, por exemplo,

$$\begin{matrix} A & I & B \\ (m \times n) & (n \times n) & (n \times p) \end{matrix} = (AI)B = \begin{matrix} A & B \\ (m \times n) & (n \times p) \end{matrix}$$

que mostra que a presença ou a ausência de I não afeta o produto. Observe que a conformidade com a dimensão é preservada quer I apareça ou não no produto.

Um caso interessante de (4.8) ocorre quando $A = I_n$, pois temos, então

$$AI_n = (I_n)^2 = I_n$$

que diz que uma matriz identidade ao quadrado é igual a si mesma. Uma generalização desse resultado é que

$$(I_n)^k = I_n \quad (k = 1, 2, \dots)$$

Uma matriz identidade permanece imutável quando é multiplicada por si mesma qualquer número de vezes. Qualquer matriz que tenha essa propriedade (a saber, $AA = A$) é denominada uma *matriz idempotente*.

Matrizes nulas

Exatamente como uma matriz identidade I faz o papel do número 1, uma *matriz nula* – ou *matriz zero* – denotada por 0 , faz o papel do número 0. Uma matriz nula é, simplesmente, uma matriz na qual todos os elementos são zero. Diferentemente de I , a matriz zero não está restrita a ser quadrada. Assim, é possível escrever

$$\underset{(2 \times 2)}{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{e} \quad \underset{(2 \times 3)}{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

e assim por diante. Uma matriz quadrada nula é idempotente, mas uma não quadrada não é. (Por quê?)

Assim como o número 0, matrizes nulas obedecem às seguintes regras de operação (sujeitas às condições de conformidade) em relação à adição e à multiplicação:

$$\begin{aligned} \underset{(m \times n)}{A} + \underset{(m \times n)}{0} &= \underset{(m \times n)}{0} + \underset{(m \times n)}{A} = \underset{(m \times n)}{A} \\ \underset{(m \times n)}{A} \underset{(n \times p)}{0} &= \underset{(m \times p)}{0} \quad \text{e} \quad \underset{(q \times m)}{0} \underset{(m \times n)}{A} = \underset{(q \times n)}{0} \end{aligned}$$

Note que, na multiplicação, a matriz nula à esquerda do sinal de igual e a matriz à direita podem ter dimensões diferentes.

EXEMPLO 2

$$A + 0 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A$$

EXEMPLO 3

$$\underset{(2 \times 3)}{A} \underset{(3 \times 1)}{0} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \underset{(2 \times 1)}{0}$$

À esquerda, a matriz nula é um vetor nulo 3×1 ; à direita, é um vetor nulo 2×1 .

Idiossincrasias da álgebra matricial

A despeito das aparentes similaridades entre a álgebra matricial e a álgebra escalar, o caso das matrizes realmente apresenta certas idiossincrasias que nos servem de alerta quanto a “tomar emprestado” da álgebra escalar sem muito questionamento. Já vimos que, em geral, $AB \neq BA$ na álgebra matricial. Agora vamos examinar mais duas dessas idiossincrasias.

A primeira é que, quando se trata de escalares, a equação $ab = 0$ sempre implica que ou a ou b é zero, mas isso não acontece na multiplicação de matrizes. Assim, temos

$$AB = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

embora nem A nem B sejam, em si, uma matriz zero.

Como outra ilustração, para escalares, a equação $cd = ce$ (com $c \neq 0$) implica que $d = e$. O mesmo não é válido para matrizes. Assim, dadas,

$$C = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \quad E = \begin{bmatrix} -2 & 1 \\ 3 & 2 \end{bmatrix}$$

verificamos que

$$CD = CE = \begin{bmatrix} 5 & 8 \\ 15 & 24 \end{bmatrix}$$

mesmo que $D \neq E$.

Na verdade, esses resultados estranhos concernem apenas à classe especial de matrizes conhecidas como *matrizes singulares*, das quais as matrizes A , B e C são exemplos. (*Grosso modo*, essas matrizes contêm uma linha que é múltipla de uma outra linha.) Não obstante, esses exemplos mostram as armadilhas propostas por estender, de maneira não justificada, teoremas algébricos a operações com matrizes.

EXERCÍCIO 4.5

Dados $A = \begin{bmatrix} -1 & 5 & 7 \\ 0 & -2 & 4 \end{bmatrix}$, $b = \begin{bmatrix} 9 \\ 6 \\ 0 \end{bmatrix}$ e $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$:

- Calcule: (a) AI (b) IA (c) Ix (d) $x'I$
Indique a dimensão da matriz identidade usada em cada caso.
- Calcule: (a) Ab (b) AIb (c) $x'I A$ (d) $x'A$
A inserção de I em (b) afeta o resultado em (a)? A exclusão de I em (d) afeta o resultado em (c)?
- Qual é a dimensão da matriz nula resultante de cada um dos seguintes?
(a) Pré-multiplicar A por uma matriz nula 5×2 .
(b) Pós-multiplicar A por uma matriz nula 3×6 .
(c) Pré-multiplicar b por uma matriz nula 2×3 .
(d) Pós-multiplicar x por uma matriz nula 1×5 .
- Mostre que a matriz diagonal

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

pode ser idempotente apenas se cada elemento da diagonal for ou 1 ou 0. Quantas matrizes diagonais numéricas idempotentes de dimensão $n \times n$ podem ser construídas, ao todo, a partir de tal matriz?

4.6 Transpostas e inversas

Quando as linhas e colunas de uma matriz A são intercambiadas – de modo que sua primeira linha é igual a sua primeira coluna e vice-versa – obtemos a *transposta* de A , que é denotada por A' ou A^T . O símbolo com “linha” (aspa simples) não é, de modo algum, novidade para nós; foi usado anteriormente para distinguir um vetor linha de um vetor coluna. Na nova terminologia que acabamos de apresentar, um vetor linha x' constitui o transposto do vetor coluna x . O índice T no símbolo alternativo é, obviamente, a notação abreviada para a palavra transposta.

EXEMPLO 1 Dados $A = \begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix}$ e $B = \begin{bmatrix} 3 & 4 \\ 1 & 7 \end{bmatrix}$, podemos intercambiar as linhas e colunas e escrever

$$A' = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} \text{ e } B' = \begin{bmatrix} 3 & 1 \\ 4 & 7 \end{bmatrix}$$

Por definição, se uma matriz A é $m \times n$, então sua transposta A' deve ser $n \times m$. Uma matriz quadrada $n \times n$, entretanto, possui uma transposta com a mesma dimensão.

EXEMPLO 2 Se $C = \begin{bmatrix} 9 & -1 \\ 2 & 0 \end{bmatrix}$ e $D = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 3 & 7 \\ 4 & 7 & 2 \end{bmatrix}$, então

$$C' = \begin{bmatrix} 9 & 2 \\ -1 & 0 \end{bmatrix} \quad \text{e} \quad D' = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 3 & 7 \\ 4 & 7 & 2 \end{bmatrix}$$

Aqui, a dimensão de cada transposta é idêntica à da matriz original.

Também observamos, em D' , um resultado notável: D' herda não somente a dimensão de D mas também a disposição original de elementos! O fato de $D' = D$ é o resultado da simetria dos elementos com referência à diagonal principal. Considerando a diagonal principal de D como um espelho, os elementos localizados à sua direita são imagens exatas dos elementos à sua esquerda; por conseguinte, a leitura da primeira linha é idêntica à da primeira coluna e assim por diante. A matriz D exemplifica a classe especial de matrizes quadradas conhecidas como *matrizes simétricas*. Um outro exemplo desse tipo de matriz é a matriz identidade I , que, como uma matriz simétrica, tem a transposta $I' = I$.

Propriedades das transpostas

As seguintes propriedades caracterizam as transpostas:

$$(A')' = A \quad (4.9)$$

$$(A + B)' = A' + B' \quad (4.10)$$

$$(AB)' = B'A' \quad (4.11)$$

A primeira afirma que a transposta da transposta é a matriz original – uma conclusão bem evidente por si só.

A segunda propriedade pode ser enunciada verbalmente assim: a transposta de uma soma é a soma das transpostas.

EXEMPLO 3 Se $A = \begin{bmatrix} 4 & 1 \\ 9 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 2 & 0 \\ 7 & 1 \end{bmatrix}$, então

$$(A + B)' = \begin{bmatrix} 6 & 1 \\ 16 & 1 \end{bmatrix}' = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

$$\text{e} \quad A' + B' = \begin{bmatrix} 4 & 9 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 7 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 1 & 1 \end{bmatrix}$$

A terceira propriedade é que a transposta de um produto é o produto das transpostas *na ordem inversa*. Para avaliar a necessidade da ordem inversa, vamos examinar a conformidade de dimensão dos dois produtos nos dois lados de (4.11). Se A for $m \times n$ e B for $n \times p$, então AB será $m \times p$, e $(AB)'$ será $p \times m$. Para que a igualdade seja válida, é necessário que a expressão do lado direito, $B'A'$, tenha dimensão idêntica. Visto que B' é $p \times n$ e A' é $n \times m$, o produto $B'A'$ é, de fato, $p \times m$, como exigido. Assim, a dimensão de $B'A'$ está correta. Note que, por outro lado, o produto $A'B'$ não é nem mesmo definido, a menos que $m = p$.

EXEMPLO 4 Dados $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ e $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$, temos

$$(AB)' = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}' = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

$$\text{e} \quad B' A' = \begin{bmatrix} 0 & 6 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 12 & 24 \\ 13 & 25 \end{bmatrix}$$

Isso verifica a propriedade.

Inversas e suas propriedades

Para uma dada matriz A , a transposta A' sempre pode ser definida. Por outro lado, sua matriz *inversa* – um outro tipo de matriz “derivada de A ” – pode ou não existir. A inversa da matriz A , denotada por A^{-1} , é definida somente se A for uma matriz quadrada, caso em que a inversa é a matriz que satisfaz a condição

$$AA^{-1} = A^{-1}A = I \quad (4.12)$$

Isto é, quer A seja pré-multiplicada ou pós-multiplicada por A^{-1} , o produto será a mesma matriz identidade. Essa é uma outra exceção à regra que diz que a multiplicação de matrizes não é comutativa.

Os seguintes pontos devem ser notados:

1. Nem toda matriz quadrada tem uma inversa – ser quadrada é uma condição *necessária*, mas *não* uma condição *suficiente* para a existência de uma inversa. Se uma matriz quadrada A tiver uma inversa, diz-se que A é *não-singular*; se A não possuir nenhuma inversa, ela é denominada uma matriz *singular*.
2. Se existir A^{-1} , então a matriz A pode ser considerada como a inversa de A^{-1} , exatamente como A^{-1} é a inversa de A . Em suma, A e A^{-1} são as inversas uma da outra.
3. Se A for $n \times n$, então A^{-1} também deve ser $n \times n$; caso contrário, não poderá ser conforme para *ambas* a pré-multiplicação e a pós-multiplicação. A matriz identidade produzida pela multiplicação também será $n \times n$.
4. Se existir uma inversa, ela será única. Para provar essa unicidade, vamos supor que verificou-se que B é uma inversa de A , de modo que

$$AB = BA = I$$

Agora, admita que há uma outra matriz C tal que $AC = CA = I$. Pré-multiplicando ambos os lados de $AB = I$ por C , verificamos que

$$CAB = CI (= C) \quad [\text{por (4.8)}]$$

Visto que $CA = I$ por premissa, a equação precedente é redutível a

$$IB = C \quad \text{ou} \quad B = C$$

Isto é, B e C devem ser uma única e a mesma matriz inversa. Por essa razão, podemos falar *da* (e não *de uma*) inversa de A .

5. Na verdade, cada uma das duas partes da condição (4.12) – a saber, $AA^{-1} = I$ e $A^{-1}A = I$ – implica a outra, de modo que satisfazer qualquer uma das equações é suficiente para estabelecer a relação inversa entre A e A^{-1} . Para provar isso, devemos mostrar que, se $AA^{-1} = I$, e se houver uma matriz B tal que $BA = I$, então $B = A^{-1}$ (de modo que $BA = I$ deve ser, realmente, a equação $A^{-1}A = I$). Vamos pós-multiplicar ambos os lados da equação dada, $BA = I$, por A^{-1} ; então

$$(BA)A^{-1} = IA^{-1}$$

$$B(AA^{-1}) = IA^{-1} \quad [\text{lei associativa}]$$

$$BI = IA^{-1} \quad [AA^{-1} = I \text{ por premissa}]$$

Portanto, como exigido,

$$B = A^{-1} \quad [\text{por 4.8)]}$$

Analogamente, pode-se demonstrar que, se $A^{-1}A = I$, então a única matriz C que resulta em $CA^{-1} = I$ é $C = A$.

EXEMPLO 5

Seja $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ e $B = \frac{1}{6} \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}$; então, visto que a posição do multiplicador escalar ($\frac{1}{6}$) em B pode ser mudada, (lei comutativa), podemos escrever

$$AB = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix} \frac{1}{6} = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} \frac{1}{6} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Isso estabelece B como a inversa de A e vice-versa. A multiplicação invertida, como esperado, também resulta na mesma matriz identidade:

$$BA = \frac{1}{6} \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

As três propriedades seguintes das matrizes inversas são de interesse. Se A e B são matrizes não-singulares de dimensão $n \times n$, então

$$(A^{-1})^{-1} = A \quad (4.13)$$

$$(AB)^{-1} = B^{-1}A^{-1} \quad (4.14)$$

$$(A')^{-1} = (A^{-1})' \quad (4.15)$$

A primeira diz que a inversa de uma inversa é a matriz original. A segunda declara que a inversa de um produto é o produto das inversas *na ordem inversa*. E a última significa que a inversa da transposta é a transposta da inversa. Note que, nessas afirmações, a existência das inversas e a satisfação da condição de conformidade são pressupostas.

A validade de (4.13) é bastante óbvia, mas vamos demonstrar (4.14) e (4.15). Dado o produto AB , vamos achar sua inversa – denominada C . De (4.12), sabemos que $CAB = I$; assim, a pós-multiplicação de ambos os lados por $B^{-1}A^{-1}$ resultará em

$$CABB^{-1}A^{-1} = IB^{-1}A^{-1} (= B^{-1}A^{-1}) \quad (4.16)$$

Mas o lado esquerdo é redutível a

$$\begin{aligned} CA(BB^{-1})A^{-1} &= CAIA^{-1} && [\text{por (4.12)}] \\ &= CAA^{-1} = CI = C && [\text{por (4.12) e (4.8)}] \end{aligned}$$

A substituição dessas em (4.16) nos diz que $C = B^{-1}A^{-1}$ ou, em outras palavras, que a inversa de AB é igual a $B^{-1}A^{-1}$, como mencionado. Nessa demonstração, a equação $AA^{-1} = A^{-1}A = I$ foi utilizada duas vezes. Note que a aplicação dessa equação é permissível se e somente se uma matriz e sua inversa forem estritamente adjacentes uma à outra em um produto. Podemos escrever $A A^{-1}B = IB = B$, mas *nunca* $A B A^{-1} = B$.

A demonstração de (4.15) é a seguinte. Dado A' , vamos achar sua inversa – denominada D . Por definição, então temos $D A' = I$. Mas sabemos que

$$(AA^{-1})' = I' = I$$

produz a mesma matriz identidade. Assim, podemos escrever

$$\begin{aligned} DA' &= (AA^{-1})' \\ &= (A^{-1})'A' \quad [\text{por (4.11)}] \end{aligned}$$

Pós-multiplicando ambos os lados por $(A')^{-1}$, obtemos

$$DA'(A')^{-1} = (A^{-1})'A'(A')^{-1}$$

$$\text{ou} \quad D = (A^{-1})' \quad [\text{por (4.12)}]$$

Assim, a inversa de A' é igual a $(A^{-1})'$, como mencionado.

Nas demonstrações que acabamos de apresentar, as operações matemáticas foram executadas em blocos inteiros de números. Se esses blocos de números não tivessem sido tratados como entidades matemáticas (matrizes), as mesmas operações teriam sido muito mais longas e complicadas. A beleza da álgebra matricial está exatamente na simplificação que proporciona a tais operações.

Matriz inversa e solução de sistema de equações lineares

A aplicação do conceito de matriz inversa à solução de um sistema de equações simultâneas é imediata e direta. Com referência ao sistema de equações em (4.3), destacamos anteriormente que ele pode ser escrito em notação matricial como

$$\underset{(3 \times 3)}{A} \underset{(3 \times 1)}{x} = \underset{(3 \times 1)}{d} \quad (4.17)$$

onde A , x e d são como definidos em (4.4). Agora, se existir a matriz inversa A^{-1} , a pré-multiplicação de ambos os lados da equação (4.17) por A^{-1} resultará em

$$A^{-1}Ax = A^{-1}d$$

ou

$$\underset{(3 \times 1)}{x} = \underset{(3 \times 3)}{A^{-1}} \underset{(3 \times 1)}{d} \quad (4.18)$$

O lado esquerdo de (4.18) é um vetor coluna de variáveis, ao passo que o produto da direita é um vetor coluna de certos números conhecidos. Assim, pela definição da igualdade de matrizes ou vetores, (4.18) mostra o conjunto de valores das variáveis que satisfazem o sistema de equações, isto é, os valores da solução. Além disso, visto que, se existir A^{-1} , ela é única, $A^{-1}d$ deve ser um vetor único de valores da solução. Por conseguinte, escreveremos o vetor x em (4.18) como x^* , para indicar seu status de solução (única).

Métodos para testar a existência da inversa e de seu cálculo serão discutidos no Capítulo 5. Aqui, contudo, podemos afirmar que a inversa da matriz A em (4.4) é

$$A^{-1} = \frac{1}{52} \begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix}$$

Assim, (4.18) resultará em

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \frac{1}{52} \begin{bmatrix} 18 & -16 & -10 \\ -13 & 26 & 13 \\ -17 & 18 & 21 \end{bmatrix} \begin{bmatrix} 22 \\ 12 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

o que resulta na solução: $x_1^* = 2$, $x_2^* = 3$ e $x_3^* = 1$.

A conclusão é que um modo de achar a solução de um sistema de equações lineares $Ax = d$, onde a matriz de coeficientes A é não-singular, é, em primeiro lugar, achar a inversa A^{-1} e pós-multiplicar A^{-1} pelo vetor de constantes d . O produto $A^{-1}d$ então resultará nos valores da solução das variáveis.

EXEMPLO 6 Como mostra o Exemplo 11 da Seção 4.2, o modelo simples de renda nacional

$$Y = C + I_0 + G_0$$

$$C = a + bY$$

pode ser escrito em notação matricial como $Ax = d$, onde

$$A = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix} \quad x = \begin{bmatrix} Y \\ C \end{bmatrix} \quad \text{e} \quad d = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$$

A inversa da matriz A é (veja explicação na Seção 5.6)

$$A^{-1} = \frac{1}{1-b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$$

Assim, a solução do modelo é $x^* = A^{-1}d$, ou

$$\begin{bmatrix} Y^* \\ C^* \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix} \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} I_0 + G_0 + a \\ b(I_0 + G_0) + a \end{bmatrix}$$

EXERCÍCIO 4.6

- Dados $A = \begin{bmatrix} 0 & 4 \\ -1 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -8 \\ 0 & 1 \end{bmatrix}$ e $C = \begin{bmatrix} 1 & 0 & 9 \\ 6 & 1 & 1 \end{bmatrix}$, ache A' , B' e C' .
- Use as matrizes dadas no Problema 1 para verificar que
 - $(A + B)' = A' + B'$
 - $(AC)' = C' A'$
- Generalize o resultado (4.11) para o caso de um produto de três matrizes provando que, para quaisquer matrizes conformes, A , B e C , a equação $(ABC)' = C' B' A'$ vale.
- Dadas as quatro matrizes seguintes, teste se qualquer uma delas é a inversa de uma outra:

$$D = \begin{bmatrix} 1 & 12 \\ 0 & 3 \end{bmatrix} \quad E = \begin{bmatrix} 1 & 1 \\ 6 & 8 \end{bmatrix} \quad F = \begin{bmatrix} 1 & -4 \\ 0 & \frac{1}{3} \end{bmatrix} \quad G = \begin{bmatrix} 4 & -\frac{1}{2} \\ -3 & \frac{1}{2} \end{bmatrix}$$

- Generalize o resultado (4.14) provando que, para quaisquer matrizes conformes não-singulares, A , B e C , a equação $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ é válida.
- Seja $A = I - X(X'X)^{-1}X'$.
 - A deve ser quadrada? $(X'X)$ deve ser quadrada? X deve ser quadrada?
 - Mostre que a matriz A é idempotente. [Nota: se X' e X não forem quadradas, é inapropriado aplicar (4.14).]

4.7 Cadeias finitas de Markov

Uma aplicação comum da álgebra matricial é encontrada nos chamados processos ou cadeias de Markov. *Processos de Markov* são usados para medir ou estimar mudanças ao longo do tempo, o

que envolve a utilização de uma matriz de transição de Markov, na qual cada valor na matriz de transição é uma probabilidade de mudar de um estado (localização, emprego etc.) para um outro estado. Há, também, um vetor que contém a distribuição inicial por todos os vários estados. Multiplicando repetidamente esse vetor pela matriz de transição, pode-se estimar mudanças nos estados ao longo do tempo.

Considere o problema da rotatividade interna dos empregados de uma empresa que tem muitas filiais ou lojas diferentes.[†] Uma ilustração simples usando duas filiais, tais como Abbotsford e Burnaby, ajudará a demonstrar o básico de um processo de Markov. Para determinar o número de empregados em Abbotsford amanhã, tomamos a probabilidade de que esses empregados permanecerão na filial de Abbotsford multiplicada pelo número total de empregados que estão atualmente em Abbotsford, o que resulta no número total de empregados que estão atualmente em Abbotsford e que continuarão lá amanhã. Adiciona-se a esse número o número de empregados da filial de Burnaby que serão transferidos para Abbotsford. Acha-se esse número multiplicando-se o número total de empregados atuais em Burnaby pela probabilidade de um empregado dessa filial ser transferido para Abbotsford. De modo semelhante, o processo seria o mesmo para determinar o número de empregados na região de Burnaby amanhã, formado pelos empregados de Burnaby que preferiram ficar e pelos empregados de Abbotsford que foram transferidos para a região de Burnaby hoje. O processo descrito envolve quatro probabilidades. Essas quatro probabilidades juntas podem ser dispostas em uma matriz, que é conhecida como uma matriz de transição de Markov (ou, simplesmente, uma “Markov”).

Denote por A_t e B_t as populações de Abbotsford e Burnaby, respectivamente, em algum momento, t . Além disso, defina as probabilidades transicionais como segue:

$P_{AA} \equiv$ probabilidade de que um A atual permaneça um A

$P_{AB} \equiv$ probabilidade de que um A atual passe para B

$P_{BB} \equiv$ probabilidade de que um B atual permaneça um B

$P_{BA} \equiv$ probabilidade de que um B atual passe para A

Se denotarmos a distribuição de empregados pelas localizações no momento t como um vetor

$$x'_t = [A_t \quad B_t]$$

e as probabilidades transicionais em forma de matriz

$$M = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}$$

então, a distribuição de empregados pelas localizações no próximo período $(t+1)$ é

$$\begin{aligned} \underset{(1 \times 2)}{x'_t} \underset{(2 \times 2)}{M} &= \underset{(1 \times 2)}{x'_{t+1}} \\ [A_t \quad B_t] \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} &= [(A_t P_{AA} + B_t P_{BA}) \quad (A_t P_{AB} + B_t P_{BB})] \\ &= [A_{t+1} \quad B_{t+1}] \end{aligned}$$

Para achar a distribuição de empregados após dois períodos

$$[A_{t+1} \quad B_{t+1}] \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} = [A_{t+2} \quad B_{t+2}]$$

[†] Gostaríamos de agradecer a Sarah Dunn por este exemplo. Esse trabalho se origina do seu projeto final como estudante do British Columbia Institute of Technology, Burnaby, BC, Canadá (junho, 2003).

$$\begin{aligned}
 [A_t \ B_t] \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} &= [A_{t+2} \ B_{t+2}] \\
 [A_t \ B_t] \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}^2 &= [A_{t+2} \ B_{t+2}]
 \end{aligned}$$

Em geral, para n períodos

$$[A_t \ B_t] \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}^n = [A_{t+n} \ B_{t+n}]$$

A matriz de probabilidade 2×2 , M , é conhecida como a *matriz de transição de Markov*. Para o caso em que n é exógeno, o processo é conhecido como uma cadeia de Markov finita.

EXEMPLO 1 Suponha que a distribuição inicial de empregados pelas duas localizações no momento $t = 0$ é

$$x'_0 = [A_0 \ B_0] = [100 \ 100]$$

Em outras palavras, inicialmente há um número igual de empregados em cada localização. Além disso, sejam as seguintes as probabilidades em forma matricial:

$$M = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}$$

Então, a distribuição de empregados pelas localizações no próximo período ($t = 1$) é

$$[100 \ 100] \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} = [110 \ 90] = [A_1 \ B_1]$$

A distribuição após dois períodos é dada por

$$\begin{aligned}
 [100 \ 100] \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}^2 &= [100 \ 100] \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix} \\
 &= [113 \ 87] = [A_2 \ B_2]
 \end{aligned}$$

A distribuição após 10 períodos ($t = 10$) é dada por

$$\begin{aligned}
 [100 \ 100] \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}^{10} &= [100 \ 100] \begin{bmatrix} 0,5174 & 0,4286 \\ 0,5174 & 0,4286 \end{bmatrix} \\
 &= [114,3 \ 85,7] = [A_{10} \ B_{10}]
 \end{aligned}$$

Observe o que acontece quando a matriz de transição de Markov é elevada a potências cada vez mais altas. A nova matriz de transição encontrada elevando a matriz original a potências cada vez maiores converge para uma matriz na qual as linhas são idênticas. Isso é denominado *estado estável*. Na sua opinião, como seria o décimo primeiro período de distribuição ou períodos de distribuição mais altos?

Caso especial: cadeias de Markov absorventes

Agora, vamos ampliar o modelo adicionando uma terceira opção: empregados podem sair da empresa, com

$P_{AE} \equiv$ probabilidade de que um A atual prefira sair (E)

$P_{BE} \equiv$ probabilidade de que um B atual prefira sair (E)

Nesse ponto, faremos as seguinte premissas adicionais:

$$P_{EA} = 0 \quad P_{EB} = 0 \quad P_{EE} = 1$$

onde P_{EA} , P_{EB} e P_{EE} são as probabilidades de que um empregado que está atualmente em E irá para A , B ou E , respectivamente. Em outras palavras, ninguém que sai da empresa volta. Também fica implícito, por essas restrições, que nossa empresa nunca substitui empregados que saem (não há novas contratações).

Começando no momento $t = 0$, nossa cadeia de Markov agora se torna

$$\begin{bmatrix} A_0 & B_0 & E_0 \end{bmatrix} \begin{bmatrix} P_{AA} & P_{AB} & P_{AE} \\ P_{BA} & P_{BB} & P_{BE} \\ P_{EA} & P_{EB} & P_{EE} \end{bmatrix}^n = \begin{bmatrix} A_n & B_n & E_n \end{bmatrix}$$

$$\begin{bmatrix} A_0 & B_0 & E_0 \end{bmatrix} \begin{bmatrix} P_{AA} & P_{AB} & P_{AE} \\ P_{BA} & P_{BB} & P_{BE} \\ 0 & 0 & 1 \end{bmatrix}^n = \begin{bmatrix} A_n & B_n & E_n \end{bmatrix}$$

(Suponha que $E_0 = 0$.)

Esse tipo de processo de Markov é conhecido como uma *cadeia de Markov absorvente*. Por causa dos valores das probabilidades transicionais encontrados na terceira linha, vemos que, uma vez que um empregado se torna um E em um estado (período de tempo), ele permanecerá ali durante todos os estados futuros (períodos de tempo). À medida que n tende ao infinito, A_n e B_n se aproximarão de zero e E_n tenderá ao valor do número total de trabalhadores no momento zero (isto é, $A_0 + B_0 + E_0$).

EXERCÍCIO 4.7

1. Considere a situação de uma demissão em massa (isto é, o fechamento de uma fábrica), na qual 1.200 pessoas ficam desempregadas e começam a procurar emprego. Nesse caso, há dois estados: empregado (E) e desempregado (U) com um vetor inicial

$$x'_0 = [E \quad U] = [0 \quad 1.200]$$

Suponha que, em qualquer período dado, uma pessoa desempregada conseguirá um emprego com probabilidade 0,7 e que, portanto, ficará desempregado com uma probabilidade de 0,3. Além disso, pessoas que estão empregadas em qualquer período de tempo dado podem perder seus empregos com uma probabilidade de 0,1 (e terão uma probabilidade de 0,9 de continuarem empregados).

- (a) Monte a matriz de transição de Markov para esse problema.
- (b) Qual será o número de desempregados após (i) 2 períodos; (ii) 3 períodos; (iii) 5 períodos; (iv) 10 períodos?
- (c) Qual é o nível de desemprego no estado estável?

CAPÍTULO 5

Modelos lineares e álgebra matricial

(CONTINUAÇÃO)

No Capítulo 4, foi demonstrado que um sistema de equações lineares, seja qual for seu tamanho, pode ser escrito em notação matricial compacta. Além disso, tal sistema de equações pode ser resolvido achando-se a inversa da matriz de coeficientes, contanto que essa inversa exista. Agora, devemos nos voltar para as questões de como testar a existência da inversa e de como achar essa inversa. Somente após termos respondido a essas perguntas será possível aplicar álgebra matricial de modo significativo a modelos econômicos.

5.1 Condições para a invertibilidade de uma matriz

Uma dada matriz de coeficientes A pode ter uma inversa (isto é, pode ser “invertível”) somente se ela for quadrada. Como salientamos anteriormente, contudo, essa condição é necessária mas não suficiente para a existência da inversa A^{-1} . Não obstante, uma matriz pode ser quadrada, mas singular (não ter uma inversa).

Condições necessárias versus condições suficientes

Os conceitos de “condição necessária” e “condição suficiente” são utilizados com frequência na economia. É importante entender os significados exatos antes de ir adiante.

Uma condição necessária tem as mesmas características de um pré-requisito: suponha que uma declaração p é verdadeira *somente se* uma outra declaração q for verdadeira; então, q constitui uma condição necessária de p . Esse enunciado é expresso simbolicamente da seguinte maneira:

$$p \Rightarrow q \quad (5.1)$$

que se lê “ p somente se q ” ou, alternativamente, “se p , então q .” Também é logicamente correto interpretar o significado de (5.1) como “ p implica q ”. É claro que pode acontecer que também tenhamos $p \Rightarrow w$ ao mesmo tempo. Então, ambas, q e w , são condições necessárias para p .

EXEMPLO 1

Se p for a declaração “uma pessoa é um pai” e q for a declaração “uma pessoa é do sexo masculino”, então a declaração lógica $p \Rightarrow q$ se aplica. Uma pessoa é um pai *somente se* for do sexo masculino, e ser do sexo masculino é uma condição necessária para a paternidade. Note, no entanto, que a recíproca não é verdadeira: a paternidade não é uma condição necessária para alguém ser do sexo masculino.

Um tipo diferente de situação é aquela em que a declaração p é verdadeira se q for verdadeira, mas p também pode ser verdadeira quando q não for verdadeira. Nesse caso, diz-se que q é uma condição suficiente para p . O fato de q ser verdadeira é suficiente para estabelecer a verdade de p , mas não é uma condição necessária para p . Esse caso é expresso simbolicamente por

$$p \Leftarrow q \quad (5.2)$$

que se lê “ p se q ” (sem a palavra *somente*) – ou, alternativamente, “se q , então p ”, como se estivéssemos lendo (5.2) ao contrário. Também podemos interpretar essa expressão como “ q implica p ”.

EXEMPLO 2 Se p for a declaração “podemos ir para a Europa” e q a declaração “pegamos um avião para a Europa”, então $p \Leftarrow q$. Pegar um avião pode ser um meio de ir para a Europa, porém, visto que o transporte por mar também é viável, pegar um avião não é um pré-requisito. Podemos escrever $p \Leftarrow q$, mas não $p \Rightarrow q$.

Em uma terceira situação possível, q é necessária e *também* suficiente para p . Nesse caso, escrevemos

$$p \Leftrightarrow q \quad (5.3)$$

que se lê “ p se e somente se q ” (também escrito como “ p iff q ”, em inglês). A seta de duas pontas é, na realidade, uma combinação dos dois tipos de setas em (5.1) e (5.2), daí a utilização conjunta dos dois termos “se” e “somente se”. Note que (5.3) determina não somente que p implica q , mas também que q implica p .

EXEMPLO 3 Se p for a declaração “há menos do que 30 dias em um mês” e q a declaração “é o mês de fevereiro”, então $p \Leftrightarrow q$. Para que um mês tenha menos do que 30 dias, é necessário que esse mês seja fevereiro. Reciprocamente, especificar fevereiro é suficiente para estabelecer que há menos do que 30 dias em um mês. Assim, q é uma condição necessária e suficiente para p .

Para provar que $p \Rightarrow q$, é preciso mostrar que q decorre logicamente de p . De modo semelhante, provar que $p \Leftarrow q$ requer demonstrar que p decorre logicamente de q . Mas, para provar que $p \Leftrightarrow q$, é preciso demonstrar que p e q decorrem uma da outra.

Condições necessárias e condições suficientes são importantes como recursos de filtragem. Considere um conjunto de candidatos a bolsas de estudo ou a cargos em uma empresa. Já que condições necessárias têm as mesmas características de pré-requisitos, elas servem para separar os candidatos em dois grupos: os que não cumprem as condições necessárias são automaticamente desqualificados; os que cumprem as condições necessárias continuam como candidatos aceitáveis. Entretanto, continuar como aceitável não é nenhuma garantia de que o candidato venha a ser bem-sucedido. Assim, condições necessárias são mais conclusivas para filtrar candidatos inadequados do que para identificar candidatos bem-sucedidos. Em geral, devemos ter sempre em mente que condições necessárias *não são*, por si sós, *suficientes*.

Condições suficientes, ao contrário de condições necessárias, servem para identificar diretamente candidatos bem-sucedidos. Um candidato que satisfaça uma condição suficiente é, automaticamente, um candidato bem-sucedido. Exatamente como condições necessárias não são suficientes por si sós, condições suficientes não são, por si sós, necessárias. Isso porque, junto com qualquer condição suficiente dada, podem existir outras condições suficientes, menos restritivas, e o candidato que não satisfizer a condição suficiente dada ainda pode se qualificar por uma condição suficiente mais fácil. Por exemplo, uma nota 10 é suficiente para ser aprovado em um curso, mas não é uma condição necessária, já que uma nota 8 também é suficiente.

O recurso de filtragem mais eficiente é encontrado nas condições necessárias e suficientes. Se um candidato não satisfizer tal condição, significa que ele está definitivamente fora e, se um candidato satisfizer essa condição, significa que ele está definitivamente dentro. Podemos encontrar uma aplicação imediata desse conceito em nossa presente discussão de invertibilidade de uma matriz.

Condições para a invertibilidade

Após cumprida a condição de ser quadrada (uma condição necessária), uma condição suficiente para a invertibilidade de uma matriz é que suas linhas sejam linearmente independentes (ou, o que quer dizer a mesma coisa, que suas *colunas* sejam linearmente independentes). Quando tomadas em conjunto, as duas condições – de ser quadrada e de independência linear das linhas – constituem a condição necessária e suficiente para a invertibilidade (invertibilidade $\Leftrightarrow$ quadrada e independência linear das linhas).

Uma matriz de coeficientes $n \times n$, A , pode ser considerada como um conjunto ordenado de vetores linhas, isto é, como um vetor coluna cujos elementos são, eles mesmos, vetores linhas:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix}$$

onde $v'_i = [a_{i1} \ a_{i2} \ \cdots \ a_{in}]$, $i = 1, 2, \dots, n$. Para que as linhas (vetores linhas) sejam linearmente independentes, nenhuma delas pode ser uma combinação linear das outras. Em termos mais formais, como mencionado na Seção 4.3, a independência linear de linhas requer que o único conjunto de escalares k_i que pode satisfazer a equação vetorial

$$\sum_{i=1}^n k_i v'_i = \mathbf{0}_{(1 \times n)} \quad (5.4)$$

seja $k_i = 0$ para todo i .

EXEMPLO 4 Se a matriz de coeficientes for

$$A = \begin{bmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 6 & 8 & 10 \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix}$$

então, visto que $[6 \ 8 \ 10] = 2[3 \ 4 \ 5]$, temos $v'_3 = 2v'_1 = 2v'_1 + 0v'_2$. Assim, a terceira linha pode ser expressa como uma combinação linear das duas primeiras, e as linhas *não* são linearmente independentes. Uma alternativa é escrever a equação anterior como

$$2v'_1 + 0v'_2 - v'_3 = [6 \ 8 \ 10] + [0 \ 0 \ 0] - [6 \ 8 \ 10] = [0 \ 0 \ 0]$$

Considerando que o conjunto de escalares que levou ao vetor zero de (5.4) não é $k_i = 0$ para todo i , resulta que as linhas são linearmente dependentes.

Ao contrário da propriedade de ser quadrada, a condição de independência linear das linhas normalmente não pode ser assegurada imediatamente. Assim, é preciso desenvolver um método para testar a independência linear entre linhas (ou colunas). Antes de nos dedicarmos a essa tarefa, entretanto, nossa motivação ficaria fortalecida se, antes de mais nada, entendêssemos, intuitivamente, por que a condição de independência linear vem junto com a condição de ser quadrada. Lembre-se de que, quando discutimos a contagem de equações e incógnitas na Seção 3.4, chegamos à conclusão geral que, para um sistema de equações ter uma solução única, não é suficiente que tenha o mesmo número de equações e de incógnitas. Além disso, as *equações* devem ser consistentes umas com as outras e também funcionalmente independentes umas das outras (o que significa, no presente contexto de sistemas lineares, *linearmente* independentes). Há uma vinculação razoavelmente óbvia entre o critério “o mesmo número de equações e incógnitas” e a *propriedade de ser quadrada* (ter o mesmo número de linhas e colunas) da matriz de coeficientes. O efeito do requisito da “independência linear entre as linhas” é impedir a inconsistência e a dependência linear também *entre as equações*. Portanto, tomados em conjunto, o requisito

duplo de ser quadrada e de independência de linhas da matriz de coeficientes equivale às condições para a existência de uma solução única enunciada na Seção 3.4.

Vamos ilustrar como a dependência linear *entre as linhas* da matriz de coeficientes pode causar inconsistência ou dependência linear *entre as próprias equações*. Seja o sistema de equações $Ax = d$ na forma

$$\begin{bmatrix} 10 & 4 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

onde a matriz de coeficientes A contém linhas linearmente dependentes: $v'_1 = 2v'_2$. (Note que suas colunas também são dependentes, pois a primeira é $\frac{5}{2}$ da segunda.) Não especificamos os valores dos termos constantes d_1 e d_2 , mas há somente *duas* possibilidades distintas para seus valores relativos: (1) $d_1 = 2d_2$ e (2) $d_1 \neq 2d_2$. Na primeira possibilidade – com, digamos, $d_1 = 12$ e $d_2 = 6$ –, as duas equações são consistentes, mas *linearmente dependentes* (exatamente como as duas linhas da matriz A), pois a primeira equação é meramente a segunda equação vezes 2. Então, uma equação é redundante, e o sistema se reduz, de fato, a uma única equação, $5x_1 + 2x_2 = 6$, com um número infinito de soluções. Na segunda possibilidade – com, digamos, $d_1 = 12$, mas $d_2 = 0$ –, as duas equações são *inconsistentes* porque, se a primeira equação ($10x_1 + 4x_2 = 12$) for verdadeira, então, dividindo cada termo por 2, podemos deduzir que $5x_1 + 2x_2 = 6$; conseqüentemente, não há possibilidade de a segunda equação ($5x_1 + 2x_2 = 0$) também ser verdadeira. Assim, não existe nenhuma solução.

A conclusão é que não há uma solução única disponível (considerando qualquer das duas possibilidades) quando as linhas da matriz de coeficientes, A , forem linearmente dependentes. Realmente, o único modo de ter uma solução única é ter linhas (ou colunas) linearmente independentes na matriz de coeficientes. Nesse caso, a matriz A será invertível, o que significa que a inversa A^{-1} existe, e que a solução única $x^* = A^{-1}d$ pode ser encontrada.

Posto de uma matriz

Mesmo que tenha sido discutido apenas em relação a matrizes quadradas, o conceito de independência das linhas aplica-se igualmente a qualquer matriz retangular $m \times n$. Se o número máximo de linhas linearmente independentes que puder ser encontrado em tal matriz for r , diz-se que o *posto* da matriz é r . (O posto também nos diz o número máximo de *colunas* linearmente independentes na dita matriz.) O posto de uma matriz $m \times n$ pode ser, no máximo, m ou n , o que for menor.

Dada uma matriz com apenas duas linhas (ou colunas), é fácil verificar a independência das linhas (ou a independência das colunas) por inspeção visual – basta verificar se uma linha (ou coluna) é o múltiplo exato da outra. Mas, no caso de uma matriz de dimensão maior, a inspeção visual pode não ser viável, e é preciso um método mais formal. Um método para determinar o posto de uma matriz A (não necessariamente quadrada), isto é, para determinar o número de linhas independentes em A , envolve transformar A em uma matriz denominada *matriz escalonada*, utilizando certas “operações elementares nas linhas”. Uma característica estrutural particular da matriz escalonada então nos dirá o posto da matriz A .

Há somente três tipos de *operações elementares nas linhas* que podem ser executadas em uma matriz:[†]

1. Troca entre quaisquer duas linhas na matriz.
2. Multiplicação (ou divisão) de uma linha por qualquer escalar $k \neq 0$.
3. Adição de “ k vezes qualquer linha” a uma outra linha.

Embora cada uma dessas operações converta uma matriz A a uma forma diferente, nenhuma delas altera o posto da matriz. É essa característica das operações elementares nas linhas que

[†] Semelhante às operações elementares nas linhas, pode haver operações elementares definidas nas colunas. Para nossas finalidades, operações nas linhas são suficientes.

nos habilita a determinar o posto de A a partir de sua matriz escalonada. O modo mais fácil de explicar o método da matriz escalonada é com um exemplo específico.

EXEMPLO 5 Encontre o posto da matriz

$$A = \begin{bmatrix} 0 & -11 & -4 \\ 2 & 6 & 2 \\ 4 & 1 & 0 \end{bmatrix}$$

a partir de sua forma escalonada. Em primeiro lugar, examinamos a primeira coluna de A para verificar a presença de elementos nulos. Se houver elementos nulos na coluna 1, transferimos esses elementos nulos para a linha de baixo da matriz. No caso de A , transferiremos o 0 (primeiro elemento da coluna 1) para a posição inferior daquela coluna, o que pode ser feito trocando a linha 1 com a linha 3 (utilizando a primeira operação elementar nas linhas). O resultado é

$$A_1 = \begin{bmatrix} 4 & 1 & 0 \\ 2 & 6 & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

Nosso próximo objetivo é remodelar a primeira coluna de A_1 para um vetor unitário e_1 como definido em (4.7). Para transformar o elemento 4 em unidade, dividimos a linha 1 de A_1 pelo escalar 4 (aplicando a segunda operação elementar nas linhas), o que dá

$$A_2 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 2 & 6 & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

Então, para transformar o elemento 2 na coluna 1 de A_2 em 0, multiplicamos a linha 1 de A_2 por -2 e, então, adicionamos o resultado à linha 2 de A_2 (aplicando a terceira operação elementar nas linhas). A matriz resultante,

$$A_3 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & 5\frac{1}{2} & 2 \\ 0 & -11 & -4 \end{bmatrix}$$

agora tem o vetor unitário desejado e_1 como sua primeira coluna. Isso feito, agora excluimos a primeira linha de A_3 de qualquer consideração adicional e continuamos a trabalhar somente com as duas linhas restantes, onde queremos criar um vetor unitário de dois elementos na segunda coluna – transformando o elemento $5\frac{1}{2}$ em 1, e o elemento -11 em 0. Para essa finalidade, precisamos dividir a linha 2 de A_3 por $5\frac{1}{2}$, transformando, desse modo, a linha no vetor $[0 \ 1 \ \frac{4}{11}]$ e, então, adicionar 11 vezes esse vetor à linha 3 de A_3 . O resultado final, na forma de

$$A_4 = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{4}{11} \\ 0 & 0 & 0 \end{bmatrix}$$

é um exemplo de matriz escalonada que, por definição, possui três características estruturais. Primeira, linhas não nulas (linhas que têm, no mínimo, um elemento não nulo) aparecem acima das linhas nulas (linhas com todos os elementos nulos). Segunda, em toda linha não nula, o primeiro elemento não nulo é a unidade. Terceira, o elemento unidade (o primeiro elemento não nulo) em qualquer linha deve aparecer à esquerda do elemento unidade contraparte da linha imediatamente seguinte. Agora já deve estar claro que todas as operações elementares nas linhas que realizamos são projetadas para produzir essas características em A_4 .

Agora, podemos simplesmente ler o posto de A a partir do número de linhas não nulas presentes na matriz escalonada A_4 . Visto que A_4 contém duas linhas não nulas, podemos concluir que $r(A) = 2$. Evidentemente, esse também é o posto das matrizes A_1 a A_4 , porque operações elementares nas linhas não alteram o posto de uma matriz.

O método de transformação de matriz escalonada se aplica tanto a matrizes que não são quadradas quanto a matrizes quadradas. Escolhemos a matriz quadrada para o Exemplo 5 porque nosso objetivo imediato é abordar a questão da invertibilidade, concernente apenas a matrizes quadradas. Por definição, para que uma matriz $n \times n$, A , seja invertível, ela deve ter n linhas (ou colunas) linearmente independentes; por consequência, deve ser de posto n , e sua matriz escalonada deve conter exatamente n linhas não nulas, sem absolutamente nenhuma linha nula. Reciprocamente, uma matriz $n \times n$ de posto n deve ser invertível. Assim, uma matriz escalonada $n \times n$ sem nenhuma linha nula deve ser invertível, como é a matriz da qual a matriz escalonada é derivada por meio de operações elementares nas linhas. No Exemplo 5, a matriz A é 3×3 , mas $r(A) = 2$; por conseguinte, A não é invertível.

EXERCÍCIO 5.1

- Nos seguintes pares de declarações, seja p a primeira declaração e q a segunda. Indique, para cada caso, se (5.1), (5.2) ou (5.3) se aplicam.
 - É um feriado; é a Proclamação da República.
 - Uma figura geométrica tem quatro lados; é um retângulo.
 - Dois pares ordenados (a, b) e (b, a) são iguais; a é igual a b .
 - Um número é racional; ele pode ser expresso como uma razão entre dois inteiros.
 - A matriz 4×4 é invertível; o posto da matriz 4×4 é 4.
 - O tanque de gasolina de meu carro está vazio; não posso dar partida em meu carro.
 - A carta é devolvida ao remetente com a indicação "endereço desconhecido"; o remetente escreveu o endereço errado no envelope.
- Seja p a declaração "uma figura geométrica é um quadrado" e sejam q as seguintes:
 - Ela tem quatro lados.
 - Ela tem quatro lados iguais.
 - Ela tem quatro lados iguais, cada um deles perpendicular ao adjacente.
 O que é verdadeiro para cada caso: $p \Rightarrow q$, $p \Leftarrow q$ ou $p \Leftrightarrow q$?
- Em cada uma das seguintes, as linhas são linearmente independentes?
 - $\begin{bmatrix} 24 & 8 \\ 9 & -3 \end{bmatrix}$
 - $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
 - $\begin{bmatrix} 0 & 4 \\ 3 & 2 \end{bmatrix}$
 - $\begin{bmatrix} -1 & 5 \\ 2 & -10 \end{bmatrix}$
- Verifique se as colunas de cada matriz no Problema 3 também são linearmente independentes. A resposta é a mesma do caso de independência das linhas?
- Encontre o posto de cada uma das matrizes seguintes a partir de sua matriz escalonada e comente a questão da invertibilidade.

$$(a) A = \begin{bmatrix} 1 & 5 & 1 \\ 0 & 3 & 9 \\ -1 & 0 & 0 \end{bmatrix}$$

$$(c) C = \begin{bmatrix} 7 & 6 & 3 & 3 \\ 0 & 1 & 2 & 1 \\ 8 & 0 & 0 & 8 \end{bmatrix}$$

$$(b) B = \begin{bmatrix} 0 & -1 & -4 \\ 3 & 1 & 2 \\ 6 & 1 & 0 \end{bmatrix}$$

$$(d) D = \begin{bmatrix} 2 & 7 & 9 & -1 \\ 1 & 1 & 0 & 1 \\ 0 & 5 & 9 & -3 \end{bmatrix}$$

- Pela definição de dependência linear entre linhas de uma matriz, uma ou mais linhas podem ser expressas como uma combinação linear de algumas outras linhas. Na matriz escalonada, a dependência linear é indicada pela presença de uma ou mais linhas nulas. O que proporciona o vínculo entre a presença de uma combinação linear de linhas em uma dada matriz e a presença de linhas nulas na matriz escalonada?

5.2 Teste de invertibilidade utilizando determinante

Para nos certificarmos que uma matriz é invertível, também podemos fazer uso do conceito de determinante.

Determinantes e invertibilidade

O determinante de uma matriz quadrada A , denotado por $|A|$, é um escalar (número) unicamente definido, associado com aquela matriz. Determinantes são definidos somente para matrizes *quadradas*. A menor matriz possível é, obviamente, a matriz 1×1 $A = [a_{11}]$. Por definição, seu determinante é igual ao próprio elemento único, a_{11} : $|A| = |a_{11}| = a_{11}$. Aqui, o símbolo $|a_{11}|$ não deve ser confundido com o símbolo parecido que indica o valor absoluto de um número. No contexto do valor absoluto, temos, por exemplo, não somente $|5| = 5$, mas também $|-5| = 5$, porque o valor absoluto de um número é seu valor numérico, sem referência ao sinal algébrico. Ao contrário, o símbolo de determinante preserva o sinal do elemento, portanto, enquanto $|8| = 8$ (um número positivo), temos $|-8| = -8$ (um número negativo). Essa distinção se mostra crucial em discussão posterior, quando aplicarmos testes em determinantes cujos resultados dependem criticamente dos sinais de determinantes de várias dimensões, incluindo os 1×1 , tais como $|a_{11}| = a_{11}$.

Para uma matriz 2×2 $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, seu determinante é definido como a soma de dois termos, como se segue:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \quad [= \text{um escalar}] \quad (5.5)$$

que é obtida multiplicando-se os dois elementos na diagonal principal de A e depois subtraindo o produto dos dois elementos restantes. Em vista da dimensão da matriz A , o determinante $|A|$ dado em (5.5) é denominado um *determinante de segunda ordem*.

EXEMPLO 1

Dados $A = \begin{bmatrix} 10 & 4 \\ 8 & 5 \end{bmatrix}$ e $B = \begin{bmatrix} 3 & 5 \\ 0 & -1 \end{bmatrix}$ seus determinantes são

$$|A| = \begin{vmatrix} 10 & 4 \\ 8 & 5 \end{vmatrix} = 10(5) - 8(4) = 18$$

$$\text{e} \quad |B| = \begin{vmatrix} 3 & 5 \\ 0 & -1 \end{vmatrix} = 3(-1) - 0(5) = -3$$

Enquanto um determinante (limitado por duas barras verticais, e não por colchetes) é, por definição, um escalar, uma matriz, como tal, não tem um valor numérico. Em outras palavras, um determinante é redutível a um número, mas uma matriz, ao contrário, é todo um bloco de números. Também deve ser enfatizado que um determinante é definido somente para uma matriz quadrada, ao passo que uma matriz não tem de ser quadrada.

Mesmo nesse estágio inicial da discussão, é possível vislumbrar a relação entre a dependência linear das linhas de uma matriz A , por um lado, e seu determinante $|A|$, pelo outro. As duas matrizes

$$C = \begin{bmatrix} c'_1 \\ c'_2 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 3 & 8 \end{bmatrix} \quad \text{e} \quad D = \begin{bmatrix} d'_1 \\ d'_2 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 8 & 24 \end{bmatrix}$$

têm ambas linhas linearmente dependentes porque $c'_1 = c'_2$ e $d'_2 = 4d'_1$. Ambos os seus determinantes também resultam iguais a zero:

$$|C| = \begin{vmatrix} 3 & 8 \\ 3 & 8 \end{vmatrix} = 3(8) - 3(8) = 0$$

$$|D| = \begin{vmatrix} 2 & 6 \\ 8 & 24 \end{vmatrix} = 2(24) - 8(6) = 0$$

Esse resultado é um forte indício de que um determinante “que desaparece” (um determinante de valor zero) pode ter algo a ver com dependência linear. Veremos que esse é, de fato, o caso. Além do mais, o valor de um determinante $|A|$ pode servir não somente como critério para testar a independência linear das linhas (e, por conseguinte, a invertibilidade) da matriz A , mas também como insumo no cálculo da inversa A^{-1} , se ela existir.

Em primeiro lugar, entretanto, precisamos ampliar nosso campo de visão discutindo determinantes de ordens mais altas.

Cálculo de um determinante de terceira ordem

Um determinante de ordem 3 é associado a uma matriz 3×3 . Dada

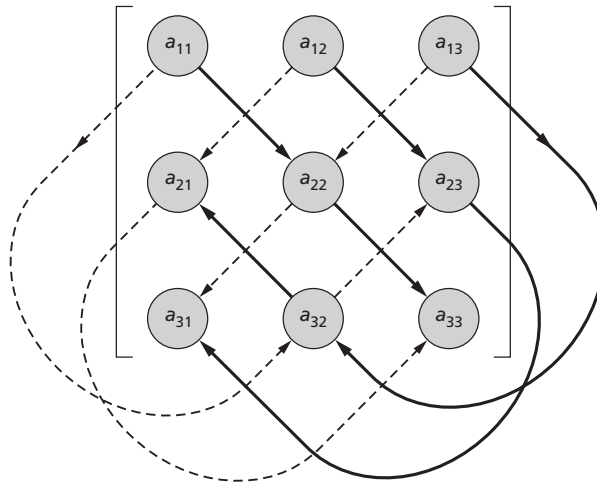
$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

seu determinante tem o valor

$$\begin{aligned} |A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \quad [= \text{um escalar}] \end{aligned} \quad (5.6)$$

Examinando em primeiro lugar a linha de baixo de (5.6), vemos o valor de $|A|$ expresso como uma soma de seis produtos de termos, três dos quais são precedidos de sinais negativos, e três de sinais positivos. Por complicada que essa soma possa parecer, não obstante há um modo muito fácil de “pegar” todos esses seis termos a partir de um dado determinante de terceira ordem. Essa operação é melhor ilustrada pelo diagrama na Figura 5.1. No determinante mostrado nessa figura, cada elemento na linha de cima foi ligado a dois outros elementos por meio de duas setas *cheias*, da seguinte maneira: $a_{11} \rightarrow a_{22} \rightarrow a_{33}$, $a_{12} \rightarrow a_{23} \rightarrow a_{31}$ e $a_{13} \rightarrow a_{32} \rightarrow a_{21}$. Cada tripla de elementos ligados dessa maneira pode ser multiplicada e seu produto considerado

FIGURA 5.1



como um dos seis termos em (5.6). Os produtos de termos resultantes das setas sólidas devem ser precedidos de sinais de mais.

Por outro lado, cada elemento da primeira linha também foi conectado com outros dois elementos por meio de duas setas *tracejadas* da seguinte maneira: $a_{11} \rightarrow a_{32} \rightarrow a_{23}$, $a_{12} \rightarrow a_{21} \rightarrow a_{33}$ e $a_{13} \rightarrow a_{22} \rightarrow a_{31}$. Cada tripla de elementos conectados dessa maneira também pode ser multiplicada e seus produtos considerados como um dos seis termos em (5.6). Esses produtos são precedidos de sinais de menos. A soma de todos os seis produtos será, então, o valor do determinante.

EXEMPLO 2

$$\begin{vmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = (2)(5)(9) + (1)(6)(7) + (3)(8)(4) - (2)(8)(6) - (1)(4)(9) - (3)(5)(7) = -9$$

EXEMPLO 3

$$\begin{vmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{vmatrix} = (-7)(1)(5) + (0)(4)(0) + (3)(6)(9) - (-7)(6)(4) - (0)(9)(5) - (3)(1)(0) = 295$$

Esse método de multiplicação pelas diagonais proporciona um modo simples de avaliar um determinante de terceira ordem, mas, infelizmente, *não* é aplicável a determinantes de ordens maiores que 3. Para estes últimos, temos de recorrer à denominada expansão de Laplace do determinante.

Cálculo de um determinante de n -ésima ordem por expansão de Laplace

Em primeiro lugar, vamos explicar o processo de *expansão de Laplace* para um determinante de terceira ordem. Voltando à primeira linha de (5.6), vemos que o valor de $|A|$ também pode ser considerado como uma soma de *três* termos, sendo cada um deles um produto entre um elemento da primeira linha e um determinante de *segunda* ordem particular. Esse último processo de avaliação de $|A|$ – por meio de certos determinantes de ordem mais baixa – ilustra a expansão de Laplace do determinante.

Os três determinantes de segunda ordem em (5.6) não são determinados arbitrariamente;

ao contrário, são especificados por meio de uma regra definida. O primeiro, $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, é um *sub-*

determinante de $|A|$ obtido pela exclusão da *primeira* linha e da *primeira* coluna de $|A|$. Esse determinante é denominado o determinante *menor* do elemento a_{11} (o elemento que se encontra na interseção da linha e da coluna excluídas) e é denotado por $|M_{11}|$. Em geral, o símbolo $|M_{ij}|$ pode ser usado para representar o determinante menor obtido pela exclusão da i -ésima linha e da j -ésima coluna de um dado determinante. Visto que um determinante menor é, em si, um determinante, ele tem um valor. Como o leitor pode verificar, os outros dois determinantes de segunda ordem em (5.6) são, respectivamente, os determinantes menores $|M_{12}|$ e $|M_{13}|$; isto é,

$$|M_{11}| \equiv \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \quad |M_{12}| \equiv \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \quad |M_{13}| \equiv \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Um conceito que guarda estreita relação com o determinante menor é o de *co-fator*. Um co-fator, denotado por $|C_{ij}|$, é um determinante menor acompanhado de um sinal algébrico definido.[†] A regra do sinal é a seguinte. Se a soma dos dois índices i e j no determinante menor $|M_{ij}|$ for par, então o co-fator adota o mesmo sinal do determinante menor; isto é, $|C_{ij}| \equiv |M_{ij}|$. Se a soma for ímpar, então, o co-fator adota o sinal oposto ao sinal do determinante menor, isto é, $|C_{ij}| \equiv -|M_{ij}|$. Em suma, temos

[†] Muitos autores usam os símbolos M_{ij} e C_{ij} (sem as barras verticais) para determinantes menores e co-fatores. Adicionamos as barras verticais para dar ênfase visual ao fato de que determinantes menores e co-fatores têm características de determinantes e, como tal, têm valores escalares.

$$|C_{ij}| \equiv (-1)^{i+j} |M_{ij}|$$

onde é óbvio que a expressão $(-1)^{i+j}$ pode ser positiva se e somente se $(i+j)$ for par. O fato de um co-fator ter um sinal específico é de extrema importância e deve ser mantido sempre em mente.

EXEMPLO 4

No determinante $\begin{vmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}$, o determinante menor do elemento 8 é

$$|M_{12}| = \begin{vmatrix} 6 & 4 \\ 3 & 1 \end{vmatrix} = -6$$

mas o co-fator desse mesmo elemento é

$$|C_{12}| = -|M_{12}| = 6$$

porque $i + j = 1 + 2 = 3$ é ímpar. De modo semelhante, o co-fator do elemento 4 é

$$|C_{23}| = -|M_{23}| = -\begin{vmatrix} 9 & 8 \\ 3 & 2 \end{vmatrix} = 6$$

Usando esses novos conceitos, podemos expressar um determinante de terceira ordem como

$$\begin{aligned} |A| &= a_{11}|M_{11}| - a_{12}|M_{12}| + a_{13}|M_{13}| \\ &= a_{11}|C_{11}| + a_{12}|C_{12}| + a_{13}|C_{13}| = \sum_{j=1}^3 a_{1j} |C_{1j}| \end{aligned} \quad (5.7)$$

isto é, como uma soma de três termos, cada um dos quais é o produto de um elemento da primeira linha e seu co-fator correspondente. Note as diferenças nos sinais dos termos $a_{12}|M_{12}|$ e $a_{12}|C_{12}|$ em (5.7). Isso ocorre porque $1 + 2$ dá um número ímpar.

A expansão de Laplace de um determinante de *terceira* ordem serve para reduzir o problema para o cálculo apenas de certos determinantes de *segunda* ordem. Obtém-se uma redução semelhante com a expansão de Laplace de determinantes de ordens mais altas que a terceira. Em um determinante de quarta ordem, $|B|$, por exemplo, a linha de cima conterá quatro elementos, $b_{11} \dots b_{14}$; assim, de acordo com o espírito de (5.7), podemos escrever

$$|B| = \sum_{j=1}^4 b_{1j} |C_{1j}|$$

onde os co-fatores $|C_{1j}|$ são de ordem 3. Então, cada co-fator de terceira ordem pode ser avaliado como em (5.6). Em geral, a expansão de Laplace de um determinante de n -ésima ordem reduzirá o problema à avaliação de n co-fatores, cada um de ordem $(n-1)$, e a aplicação repetida do processo levará metodicamente a determinantes de ordens cada vez mais baixas, culminando, por fim, nos determinantes básicos de segunda ordem definidos em (5.5). Então, o valor do determinante original pode ser calculado com facilidade.

Embora o processo de expansão de Laplace tenha sido expresso em termos dos co-fatores dos elementos da primeira linha, também é viável expandir um determinante pelo co-fator de qualquer linha ou, pelo mesmo critério, de qualquer coluna. Por exemplo, se a primeira coluna de um determinante de terceira ordem $|A|$ consistir nos elementos a_{11} , a_{21} e a_{31} , a expansão pelos co-fatores desses elementos também dará o valor de $|A|$:

$$|A| = a_{11}|C_{11}| + a_{21}|C_{21}| + a_{31}|C_{31}| = \sum_{i=1}^3 a_{i1} |C_{i1}|$$

EXEMPLO 5

Dado $|A| = \begin{vmatrix} 5 & 6 & 1 \\ 2 & 3 & 0 \\ 7 & -3 & 0 \end{vmatrix}$, a expansão pela primeira *linha* produz o resultado

$$|A| = 5 \begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 6 \begin{vmatrix} 2 & 0 \\ 7 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = 0 + 0 - 27 = -27$$

Mas a expansão pela primeira *coluna* dá a resposta idêntica:

$$|A| = 5 \begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} - 2 \begin{vmatrix} 6 & 1 \\ -3 & 0 \end{vmatrix} + 7 \begin{vmatrix} 6 & 1 \\ 3 & 0 \end{vmatrix} = 0 - 6 - 21 = -27$$

No que tange ao cálculo numérico, esse fato nos dá uma oportunidade de escolher, para a expansão, uma linha ou coluna “fácil”. Uma linha ou coluna que tenha o maior número de 0s ou 1s é sempre preferível para essa finalidade, porque 0 vezes seu co-fator é simplesmente 0, de modo que o termo será descartado, e 1 vez seu co-fator é simplesmente o próprio co-fator, de modo que ao menos uma etapa de multiplicação pode ser poupada. No Exemplo 5, o modo mais fácil de expandir o determinante é pela terceira coluna, que é composta dos elementos 1, 0 e 0. Portanto, poderíamos tê-lo avaliado assim:

$$|A| = 1 \begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = -6 - 21 = -27$$

Em resumo, o valor de um determinante $|A|$ de ordem n pode ser encontrado pela expansão de Laplace de *qualquer linha* ou de *qualquer coluna*, da seguinte maneira:

$$\begin{aligned} |A| &= \sum_{i=1}^n a_{ij} |C_{ij}| \quad [\text{expansão pela } i\text{-ésima linha}] \\ &= \sum_{i=1}^n a_{ij} |C_{ij}| \quad [\text{expansão pela } j\text{-ésima coluna}] \end{aligned} \quad (5.8)$$

EXERCÍCIO 5.2

1. Calcule os seguintes determinantes:

(a) $\begin{vmatrix} 8 & 1 & 3 \\ 4 & 0 & 1 \\ 6 & 0 & 3 \end{vmatrix}$

(c) $\begin{vmatrix} 4 & 0 & 2 \\ 6 & 0 & 3 \\ 8 & 2 & 3 \end{vmatrix}$

(e) $\begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix}$

(b) $\begin{vmatrix} 1 & 2 & 3 \\ 4 & 7 & 5 \\ 3 & 6 & 9 \end{vmatrix}$

(d) $\begin{vmatrix} 1 & 1 & 4 \\ 8 & 11 & -2 \\ 0 & 4 & 7 \end{vmatrix}$

(f) $\begin{vmatrix} x & 5 & 0 \\ 3 & y & 2 \\ 9 & -1 & 8 \end{vmatrix}$

2. Determine os sinais que deverão acompanhar os determinantes menores relevantes para obter os seguintes co-fatores de um determinante: $|C_{13}|$, $|C_{23}|$, $|C_{33}|$, $|C_{41}|$ e $|C_{34}|$.

3. Dado $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$, ache os determinantes menores e co-fatores dos elementos a , b e f .

4. Calcule os seguintes determinantes:

$$(a) \begin{vmatrix} 1 & 2 & 0 & 9 \\ 2 & 3 & 4 & 6 \\ 1 & 6 & 0 & -1 \\ 0 & -5 & 0 & 8 \end{vmatrix} \quad (b) \begin{vmatrix} 2 & 7 & 0 & 1 \\ 5 & 6 & 4 & 8 \\ 0 & 0 & 9 & 0 \\ 1 & -3 & 1 & 4 \end{vmatrix}$$

5. No primeiro determinante do Problema 4, encontre o valor do co-fator do elemento 9.

6. Encontre os determinantes menores e co-fatores da terceira linha, dado

$$A = \begin{bmatrix} 9 & 11 & 4 \\ 3 & 2 & 7 \\ 6 & 10 & 4 \end{bmatrix}$$

7. Use a expansão de Laplace para achar o determinante de

$$A = \begin{bmatrix} 15 & 7 & 9 \\ 2 & 5 & 6 \\ 9 & 0 & 12 \end{bmatrix}$$

5.3 Propriedades básicas de determinantes

Agora podemos discutir algumas propriedades de determinantes que nos permitirão “descobrir” a conexão entre a dependência linear das linhas de uma matriz quadrada e o anulamento do determinante daquela matriz.

Aqui serão discutidas cinco propriedades básicas. Essas propriedades são comuns a determinantes de todas as ordens, embora nossa ilustração aborde, quase sempre, determinantes de segunda ordem:

Propriedade I A troca de linhas por colunas não afeta o valor de um determinante. Em outras palavras, o determinante de uma matriz A tem o mesmo valor do de sua transposta A' , isto é, $|A| = |A'|$.

EXEMPLO 1

$$\begin{vmatrix} 4 & 3 \\ 5 & 6 \end{vmatrix} = \begin{vmatrix} 4 & 5 \\ 3 & 6 \end{vmatrix} = 9$$

EXEMPLO 2

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$$

Propriedade II A troca de quaisquer *duas* linhas (ou de quaisquer *duas* colunas) alterará o sinal, mas não o valor numérico do determinante. (Essa propriedade está obviamente relacionada com a primeira operação elementar nas linhas de uma matriz).

EXEMPLO 3

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \text{ mas a troca de duas linhas dá}$$

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - ad = -(ad - bc)$$

EXEMPLO 4

$$\begin{vmatrix} 0 & 1 & 3 \\ 2 & 5 & 7 \\ 3 & 0 & 1 \end{vmatrix} = -26, \text{ mas a troca da primeira com a terceira coluna dá}$$

$$\begin{vmatrix} 3 & 1 & 0 \\ 7 & 5 & 2 \\ 1 & 0 & 3 \end{vmatrix} = 26.$$

Propriedade III A multiplicação de qualquer *uma* linha (ou de qualquer *uma* coluna) por um escalar k multiplicará o valor do determinante por k . (Essa propriedade está relacionada com a segunda operação elementar nas linhas de uma matriz).

EXEMPLO 5 Multiplicando a linha de cima do determinante no Exemplo 3 por k , obtemos

$$\begin{vmatrix} ka & kb \\ c & d \end{vmatrix} = kad - kbc = k(ad - bc) = k \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

É importante distinguir entre as duas expressões kA e $k|A|$. Ao multiplicar uma *matriz* A por um escalar k , todos os elementos em A serão multiplicados por k . Mas, se lermos a equação deste exemplo da direita para a esquerda, deve ficar claro que, ao multiplicar um *determinante* $|A|$ por k , somente uma única linha (ou coluna) será multiplicada por k . Portanto, essa equação nos dá, na verdade, uma regra para fatorar um determinante: sempre que qualquer linha ou coluna contiver um divisor comum, ela pode ser fatorada para fora do determinante.

EXEMPLO 6 Fatorando a primeira coluna e a segunda linha, uma por vez, temos

$$\begin{vmatrix} 15a & 7b \\ 12c & 2d \end{vmatrix} = 3 \begin{vmatrix} 5a & 7b \\ 4c & 2d \end{vmatrix} = 3(2) \begin{vmatrix} 5a & 7b \\ 2c & d \end{vmatrix} = 6(5ad - 14bc)$$

O cálculo direto do determinante original produzirá, é claro, a mesma resposta.

Ao contrário, a fatoração de uma *matriz* requer a presença de um divisor comum para *todos* os seus elementos, como em

$$\begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix} = k \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Propriedade IV A adição (ou subtração) de um múltiplo de qualquer linha a (de) outra linha não altera o valor do determinante. O mesmo é válido se substituirmos a palavra *linha* por *coluna* na declaração anterior. (Essa propriedade está relacionada com a terceira operação elementar nas linhas de uma matriz).

EXEMPLO 7 Adicionar k vezes a linha de cima do determinante no Exemplo 3 à sua segunda linha resultará no determinante original:

$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = a(d + kb) - b(c + ka) = ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Propriedade V Se uma linha (ou coluna) for um múltiplo de qualquer outra linha (ou coluna), o valor do determinante será zero. Um caso especial dessa propriedade é que, quando duas linhas (ou duas colunas) são *idênticas*, o determinante será nulo.

EXEMPLO 8

$$\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix} = 2ab - 2ab = 0 \quad \begin{vmatrix} c & c \\ d & d \end{vmatrix} = cd - cd = 0$$

Exemplos adicionais desse tipo de determinante “que desaparece” podem ser encontrados no Exercício 5.2-1.

Essa importante propriedade é, de fato, uma consequência lógica da Propriedade IV. Para entender isso, vamos aplicar a Propriedade IV aos dois determinantes no Exemplo 8 e ver o que acontece. No primeiro determinante, tente subtrair duas vezes a segunda linha da linha de cima;

no segundo determinante, subtraia a segunda coluna da primeira coluna. Visto que essas operações não alteram os valores dos determinantes, podemos escrever

$$\begin{vmatrix} 2a & 2b \\ a & b \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ a & b \end{vmatrix} \quad \begin{vmatrix} c & c \\ d & d \end{vmatrix} = \begin{vmatrix} 0 & c \\ 0 & d \end{vmatrix}$$

Os novos determinantes (reduzidos) agora contêm, respectivamente, uma linha e uma coluna de zeros; assim, suas expansões de Laplace devem resultar em um valor zero em ambos os casos. Em geral, quando uma linha (coluna) é múltipla de uma outra linha (coluna), a aplicação da Propriedade IV sempre pode reduzir todos os elementos daquela linha (coluna) a zero e, portanto, segue a Propriedade V.

As propriedades básicas que acabamos de discutir são úteis de diversas maneiras. Uma razão é que elas podem ser de grande ajuda para simplificar a tarefa de calcular determinantes. Subtraindo múltiplos de uma linha (ou coluna) de uma outra linha (ou coluna), por exemplo, os elementos do determinante podem ser reduzidos a números muito menores e mais simples. A fatoração, se viável, também pode fazer o mesmo. Se pudermos aplicar essas propriedades para transformar alguma linha ou coluna em uma forma que contenha maioria de 0s ou 1s, a expansão de Laplace do determinante se tornará uma tarefa muito mais fácil.

Critério com determinantes para a invertibilidade

Nossa preocupação primordial agora é ligar a dependência linear de linhas com o anulamento de um determinante. Para essa finalidade, podemos invocar a Propriedade V. Considere o sistema de equações $Ax = d$:

$$\begin{bmatrix} 3 & 4 & 2 \\ 15 & 20 & 10 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Esse sistema pode ter uma solução única se e somente se as linhas da matriz de coeficientes A forem linearmente independentes, de modo que A seja invertível. Mas a segunda linha é cinco vezes a primeira; na verdade, as linhas são *dependentes* e, por conseguinte, não existe solução única. Essa dependência de linha foi detectada por inspeção visual, mas, em virtude da Propriedade V, também poderíamos tê-la descoberto pelo fato de $|A| = 0$.

É claro que a dependência das linhas em uma matriz pode assumir um padrão mais intrincado e oculto. Por exemplo, na matriz

$$B = \begin{bmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ v'_3 \end{bmatrix}$$

existe dependência entre as linhas porque $2v'_1 - v'_2 - 3v'_3 = 0$; porém, esse fato desafia a detecção visual. Contudo, mesmo nesse caso a Propriedade V nos dará um determinante que se anula, $|B| = 0$, visto que, adicionando três vezes v'_3 a v'_2 e dela subtraindo duas vezes v'_1 , a segunda linha pode ser reduzida a um vetor zero. Em geral, *qualquer* padrão de dependência linear entre linhas será refletido em um determinante que se anula – e é aí que está a beleza da Propriedade V! Reciprocamente, se as linhas forem linearmente independentes, o determinante deve ter um valor não nulo.

Nos dois parágrafos anteriores, vinculamos a invertibilidade de uma matriz principalmente à independência linear entre *linhas*. Mas, ocasionalmente, declaramos que, para uma matriz *quadrada* A , independência entre linhas $\Leftrightarrow$ independência entre colunas. Agora já temos subsídios para provar aquela declaração:

Segundo a Propriedade I, sabemos que $|A| = |A'|$. Visto que a independência entre linhas de $A \Leftrightarrow |A| \neq 0$, também podemos afirmar que independência entre linhas de $A \Leftrightarrow |A'| \neq 0$. Mas $|A'| \neq 0 \Leftrightarrow$ independência entre linhas da transposta $A' \Leftrightarrow$ independência entre colunas de A (as linhas de A' são, por definição, as colunas de A). Portanto, independência entre *linhas* de $A \Leftrightarrow$ independência entre *colunas* de A .

Agora podemos resumir nossa discussão do teste de invertibilidade. Dado um sistema de equações lineares $Ax = d$, onde A é uma matriz de coeficientes $n \times n$,

$$\begin{aligned} |A| \neq 0 &\Leftrightarrow \text{há independência entre as linhas (colunas) da matriz } A \\ &\Leftrightarrow A \text{ é invertível} \\ &\Leftrightarrow \text{existe } A^{-1} \\ &\Leftrightarrow \text{existe uma solução única } x^* = A^{-1}d \end{aligned}$$

Assim, o valor do determinante da matriz de coeficientes, $|A|$, proporciona um critério conveniente para testar a invertibilidade da matriz A e a existência de uma solução única para o sistema de equações $Ax = d$. Note, contudo, que o critério com determinantes nada diz sobre os sinais algébricos dos valores de solução; isto é, mesmo que uma solução única esteja garantida quando $|A| \neq 0$, às vezes podemos obter valores de solução negativos que são inadmissíveis em termos econômicos.

EXEMPLO 9 O sistema de equações

$$\begin{aligned} 7x_1 - 3x_2 - 3x_3 &= 7 \\ 2x_1 + 4x_2 + x_3 &= 0 \\ -2x_2 - x_3 &= 2 \end{aligned}$$

possui uma solução única? O determinante $|A|$ é

$$\begin{vmatrix} 7 & -3 & -3 \\ 2 & 4 & 1 \\ 0 & -2 & -1 \end{vmatrix} = -8 \neq 0$$

Portanto, existe uma solução única.

Redefinição do posto de uma matriz

O posto de uma matriz A foi definido anteriormente como o número máximo de linhas linearmente independentes de A . Em vista da ligação entre independência entre linhas e o valor não nulo do determinante, podemos redefinir o posto de uma matriz $m \times n$ como a ordem máxima de um determinante não nulo que pode ser construído a partir das linhas e colunas daquela matriz. O posto de qualquer matriz é um número único.

É óbvio que o posto pode ser, no máximo, m ou n , o que for menor, porque um determinante é definido somente para uma matriz quadrada e, no caso de uma matriz de dimensão 3×5 , por exemplo, os maiores determinantes possíveis (nulos ou não) serão de ordem 3. Simbolicamente, esse fato pode ser expresso da seguinte maneira:

$$r(A) \leq \min \{m, n\}$$

que se lê: “O posto de A é menor ou igual ao mínimo do conjunto dos dois números m e n ”. O posto de uma matriz invertível $n \times n$, A , tem de ser n ; nesse caso, podemos escrever $r(A) = n$.

É possível que, às vezes, estejamos interessados no posto do produto de duas matrizes. Nesse caso, a seguinte regra pode ser útil:

$$r(AB) \leq \min \{r(A), r(B)\} \quad (5.9)$$

Embora essa regra não resulte em um valor único de $r(AB)$, mesmo assim, aplicá-la pode levar a resultados únicos. Em particular, podemos usar (5.9) para mostrar que, se a matriz A , com $r(A) = j$, for multiplicada por qualquer matriz invertível B (conforme) o posto da matriz produto AB (ou BA , conforme o caso), deve ser j . Provaremos isso para o produto AB (o caso de BA é análogo). Em primeiro lugar, examinando o lado direito de (5.9), vemos apenas três casos possíveis: (i) $r(A) < r(B)$, (ii) $r(A) = r(B)$, e (iii) $r(A) > r(B)$. Para os casos (i) e (ii), (5.9) reduz diretamente a $r(AB) \leq r(A) = j$. Para o caso (iii), achamos que $r(AB) \leq r(B) < r(A) = j$. Assim, de qualquer modo, obtemos

$$r(AB) \leq r(A) = j \quad (5.10)$$

Agora considere a identidade $(AB)B^{-1} = A$. Por (5.9), podemos escrever

$$r[(AB)B^{-1}] \leq \min\{r(AB), r(B^{-1})\}$$

Aplicando o mesmo raciocínio que nos levou a (5.10), podemos concluir que

$$r[(AB)B^{-1}] \leq r(AB)$$

Visto que o lado esquerdo da expressão dessa desigualdade é igual a $r(A) = j$, podemos escrever

$$j \leq r(AB) \quad (5.11)$$

Mas (5.10) e (5.11) não podem ser satisfeitas simultaneamente a menos que $r(AB) = j$. Assim, o posto da matriz produto AB deve ser j , como afirmado.

EXERCÍCIO 5.3

- Use o determinante $\begin{vmatrix} 4 & 0 & -1 \\ 2 & 1 & -7 \\ 3 & 3 & 9 \end{vmatrix}$ para verificar as quatro primeiras propriedades de determinantes.
- Mostre que, quando todos os elementos de um determinante de n -ésima ordem $|A|$ são multiplicados por um número k , o resultado será $k^n|A|$.
- Quais propriedades de determinantes nos permitem escrever o seguinte?

$$(a) \begin{vmatrix} 9 & 18 \\ 27 & 56 \end{vmatrix} = \begin{vmatrix} 9 & 18 \\ 0 & 2 \end{vmatrix} \quad (b) \begin{vmatrix} 9 & 27 \\ 4 & 2 \end{vmatrix} = 18 \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix}$$

- Faça um teste para verificar se as matrizes seguintes são invertíveis:

$$(a) \begin{bmatrix} 4 & 0 & 1 \\ 19 & 1 & -3 \\ 7 & 1 & 0 \end{bmatrix} \quad (c) \begin{bmatrix} 7 & -1 & 0 \\ 1 & 1 & 4 \\ 13 & -3 & -4 \end{bmatrix}$$

$$(b) \begin{bmatrix} 4 & -2 & 1 \\ -5 & 6 & 0 \\ 7 & 0 & 3 \end{bmatrix} \quad (d) \begin{bmatrix} -4 & 9 & 5 \\ 3 & 0 & 1 \\ 10 & 8 & 6 \end{bmatrix}$$

- O que você conclui sobre o posto de cada matriz no Problema 4?
- Alguns dos conjuntos de vetores tridimensionais dados abaixo pode gerar abranger o espaço tridimensional? Justifique sua resposta.
 - $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 4 & 2 \end{bmatrix}$
 - $\begin{bmatrix} 8 & 1 & 3 \\ 1 & 2 & 8 \\ -7 & 1 & 5 \end{bmatrix}$

7. Escreva o modelo simples de renda nacional (3.23) no formato $Ax = d$ (sendo Y a primeira variável no vetor x) e depois faça um teste para verificar se a matriz de coeficientes A é invertível.
8. Comente a validade dos seguintes enunciados:
 - (a) "Dada qualquer matriz A , sempre podemos derivar dela uma transposta e um determinante."
 - (b) "Multiplicar por 2 cada elemento de um determinante $n \times n$ dobrará o valor daquele determinante."
 - (c) "Se uma matriz quadrada A tem determinante nulo, então podemos ter certeza de que o sistema de equações $Ax = d$ é invertível."

5.4 Encontrando a matriz inversa

Se a matriz A no sistema de equações lineares $Ax = d$ for invertível, então, existe A^{-1} , e a solução do sistema será $x^* = A^{-1}d$. Aprendemos a testar a invertibilidade de A pelo critério $|A| \neq 0$. A próxima pergunta é: como achamos a inversa A^{-1} se A passar no teste?

Expansão de um determinante por co-fatores espúrios

Antes de responder a essa pergunta, vamos discutir uma outra propriedade importante de determinantes.

Propriedade VI A expansão de um determinante por *co-fatores espúrios* (os co-fatores de uma linha ou coluna "errada") sempre dá um valor zero.

EXEMPLO 1

Se expandirmos o determinante $\begin{vmatrix} 4 & 1 & 2 \\ 5 & 2 & 1 \\ 1 & 0 & 3 \end{vmatrix}$ usando os elementos de sua *primeira* linha, mas os co-fatores dos elementos da *segunda* linha

$$|C_{21}| = - \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = -3 \quad |C_{22}| = \begin{vmatrix} 4 & 2 \\ 1 & 3 \end{vmatrix} = 10 \quad |C_{23}| = - \begin{vmatrix} 4 & 1 \\ 1 & 0 \end{vmatrix} = 1$$

obtemos $a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| = 4(-3) + 1(10) + 2(1) = 0$.

De modo mais geral, aplicar o mesmo tipo de expansão por co-fatores espúrios, como descrito no Exemplo 1, ao determinante $|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ dará uma soma zero de produtos como segue:

$$\begin{aligned} \sum_{i=1}^3 a_{1i} |C_{2i}| &= a_{11}|C_{21}| + a_{12}|C_{22}| + a_{13}|C_{23}| \\ &= -a_{11} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{13} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ &= -a_{11}a_{12}a_{33} + a_{11}a_{13}a_{32} + a_{11}a_{12}a_{33} - a_{12}a_{13}a_{31} \\ &\quad - a_{11}a_{13}a_{32} + a_{12}a_{13}a_{31} = 0 \end{aligned} \quad (5.12)$$

A razão desse resultado está no fato de a soma de produtos em (5.12) poder ser considerada como o resultado da expansão *regular* pela segunda linha de um outro determinante $|A^*| \equiv$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \text{ que é diferente de } |A| \text{ somente em sua segunda linha, e cujas duas primeiras li-}$$

nhas são idênticas. Como exercício, escreva os co-fatores das segundas linhas de $|A^*|$ e verifique se eles são exatamente os mesmos co-fatores que apareceram em (5.12) – e com os sinais corretos. Visto que $|A^*| = 0$, por causa de suas duas linhas idênticas, a expansão por co-fatores espúrios mostrada em (5.12) necessariamente também resultará em um valor zero.

A Propriedade VI é válida para determinantes de todas as ordens e se aplica quando um determinante é expandido por co-fatores espúrios de qualquer linha ou de qualquer coluna. Assim, podemos afirmar que para um determinante de ordem n vale o seguinte:

$$\begin{aligned} \sum_{j=1}^n a_{ij} |C_{ij}| &= 0 \quad (i \neq i') && \text{[expansão pela } i\text{-ésima linha e} \\ &&& \text{co-fatores da } i'\text{-ésima 'linha [} i'\text{th]} \\ \sum_{i=1}^n a_{ij} |C_{ij}| &= 0 \quad (j \neq j') && \text{[expansão pela } j\text{-ésima coluna e} \\ &&& \text{co-fatores da } j'\text{-ésima 'coluna [} j'\text{th]} \end{aligned} \quad (5.13)$$

Compare cuidadosamente (5.13) com (5.8). Na última (expansão regular de Laplace), os índices de a_{ij} e de $|C_{ij}|$ devem ser idênticos em cada termo de produto na soma. Na expansão por co-fatores espúrios, tal como em (5.13), por outro lado, um dos dois índices (um valor escolhido de i' ou j') está, inevitavelmente, “fora de lugar”.

Inversão de matriz

A Propriedade VI, como resumida em (5.13), é uma ajuda direta para desenvolver um método para a inversão de matrizes, isto é, para achar a inversa de uma matriz.

Suponha que temos uma dada matriz invertível $n \times n$ A :

$$A_{(n \times n)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (|A| \neq 0) \quad (5.14)$$

Visto que cada elemento de A tem um co-fator $|C_{ij}|$, é possível formar uma matriz de co-fatores substituindo cada elemento a_{ij} em (5.14) por seu co-fator $|C_{ij}|$. Essa matriz de co-fatores, denotada por $C = [|C_{ij}|]$, também deve ser $n \times n$. Para nossas finalidades atuais, contudo, a transposta de C é de maior interesse. Essa transposta C' é comumente denominada *adjunta* de A e é representada por $\text{adj } A$. Escrita por extenso, a adjunta toma a forma

$$C'_{(n \times n)} \equiv \text{adj } A \equiv \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \quad (5.15)$$

As matrizes A e C' são conformes para multiplicação e seu produto AC' é uma outra matriz $n \times n$ na qual cada elemento é uma soma de produtos. Utilizando a fórmula da expansão de Laplace, bem como a Propriedade VI de determinantes, o produto AC' pode ser expresso como se segue:

$$\begin{aligned}
 AC'_{(n \times n)} &= \begin{bmatrix} \sum_{j=1}^n a_{1j}|C_{1j}| & \sum_{j=1}^n a_{1j}|C_{2j}| & \cdots & \sum_{j=1}^n a_{1j}|C_{nj}| \\ \sum_{j=1}^n a_{2j}|C_{1j}| & \sum_{j=1}^n a_{2j}|C_{2j}| & \cdots & \sum_{j=1}^n a_{2j}|C_{nj}| \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^n a_{nj}|C_{1j}| & \sum_{j=1}^n a_{nj}|C_{2j}| & \cdots & \sum_{j=1}^n a_{nj}|C_{nj}| \end{bmatrix} \\
 &= \begin{bmatrix} |A| & 0 & \cdots & 0 \\ 0 & |A| & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & |A| \end{bmatrix} \quad [\text{por (5.8) e (5.13)}] \\
 &= |A| \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = |A|I_n \quad [\text{por fatoração}]
 \end{aligned}$$

Como o determinante $|A|$ é um escalar não nulo, pode-se dividir ambos os lados da equação $AC' = |A|I$ por $|A|$. O resultado é

$$\frac{AC'}{|A|} = I \quad \text{ou} \quad A \frac{C'}{|A|} = I$$

Pré-multiplicando ambos os lados da última equação por A^{-1} e usando o resultado que $A^{-1}A = I$, podemos obter $\frac{C'}{|A|} = A^{-1}$ ou

$$A^{-1} = \frac{1}{|A|} \text{adj } A \quad [\text{por (5.15)}] \quad (5.16)$$

Agora, encontramos um modo de inverter a matriz A !

O procedimento geral para encontrar a inversa de uma matriz quadrada A envolve, assim, as seguintes etapas: (1) encontrar $|A|$ [precisaremos passar para as etapas subseqüentes se e somente se $|A| \neq 0$, pois se $|A| = 0$, a inversa em (5.16) será indefinida]; (2) encontrar os co-fatores de todos os elementos de A e arranjá-los como uma matriz $C = [|C_{ij}|]$; (3) tomar a transposta de C para obter a $\text{adj } A$; e (4) dividir $\text{adj } A$ pelo determinante $|A|$. O resultado será a inversa desejada A^{-1} .

EXEMPLO 2

Encontre a inversa de $A = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$. Visto que $|A| = -2 \neq 0$, a inversa A^{-1} existe. O co-fator de cada elemento é, neste caso, um determinante 1×1 , que é definido, simplesmente, como o elemento escalar daquele determinante em si (isto é, $|a_{ji}| \equiv a_{ji}$). Assim, temos

$$C = \begin{bmatrix} |C_{11}| & |C_{12}| \\ |C_{21}| & |C_{22}| \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -2 & 3 \end{bmatrix}$$

Observe os sinais de menos ligados a 1 e 2, como é exigido para co-fatores. Transpondo a matriz de co-fatores, temos

$$\text{adj } A = \begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix}$$

portanto, a inversa A^{-1} pode ser escrita como

$$A^{-1} = \frac{1}{|A|} \text{adj } A = -\frac{1}{2} \begin{bmatrix} 0 & -2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

EXEMPLO 3

Encontre a inversa de $B = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 3 & 2 \\ 3 & 0 & 7 \end{bmatrix}$. Visto que $|B| = 99 \neq 0$, a inversa B^{-1} também existe. A matriz de co-fatores é

$$\begin{bmatrix} \begin{vmatrix} 3 & 2 \\ 0 & 7 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 3 & 7 \end{vmatrix} & \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} \\ \begin{vmatrix} 1 & -1 \\ 0 & 7 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 3 & 7 \end{vmatrix} & -\begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} \\ -\begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} 4 & 1 \\ 0 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 21 & 6 & -9 \\ -7 & 31 & 3 \\ 5 & -8 & 12 \end{bmatrix}$$

Portanto,

$$\text{adj } B = \begin{bmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

e a matriz inversa desejada é

$$B^{-1} = \frac{1}{|B|} \text{adj } B = \frac{1}{99} \begin{bmatrix} 21 & -75 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{bmatrix}$$

Você pode verificar se os resultados nos Exemplos 2 e 3 satisfazem $AA^{-1} = A^{-1}A = I$ e $BB^{-1} = B^{-1}B = I$, respectivamente.

EXERCÍCIO 5.4

- Suponha que expandamos um determinante de quarta ordem pela sua *terceira coluna* e pelos co-fatores dos elementos da *segunda coluna*. Como você escreveria a soma de produtos resultante em notação Σ ? Qual será a soma de produtos em notação Σ se a expandirmos pela *segunda linha* e pelos co-fatores dos elementos da *quarta linha*?
- Encontre a inversa de cada uma das seguintes matrizes:
 - $A = \begin{bmatrix} 5 & 2 \\ 0 & 1 \end{bmatrix}$
 - $B = \begin{bmatrix} -1 & 0 \\ 9 & 2 \end{bmatrix}$
 - $C = \begin{bmatrix} 3 & 7 \\ 3 & -1 \end{bmatrix}$
 - $D = \begin{bmatrix} 7 & 6 \\ 0 & 3 \end{bmatrix}$
- Aproveitando suas respostas para o Problema 2, formule uma regra de duas etapas para achar a adjunta de uma matriz 2×2 dada, A : na primeira etapa, indique o que deve ser feito com os elementos das duas diagonais de A para obter os elementos da diagonal de $\text{adj } A$; na segunda etapa, indique o que deve ser feito com os dois elementos que estão fora da diagonal de A . (*Atenção:* essa regra se aplica somente a matrizes 2×2 .)
 - Adicione uma terceira etapa que, juntamente com as duas etapas anteriores, resulte em uma matriz inversa 2×2 A^{-1} .
- Encontre a inversa de cada uma das seguintes matrizes:

$$(a) E = \begin{bmatrix} 4 & -2 & 1 \\ 7 & 3 & 0 \\ 2 & 0 & 1 \end{bmatrix} \quad (c) G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$(b) F = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix} \quad (d) H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

5. Encontre a inversa de

$$A = \begin{bmatrix} 4 & 1 & -5 \\ -2 & 3 & 1 \\ 3 & -1 & 4 \end{bmatrix}$$

6. Resolva o sistema $Ax = d$ por inversão de matriz, onde

$$\begin{array}{ll} \text{(a)} & 4x + 3y = 28 \\ & 2x + 5y = 42 \end{array} \quad \begin{array}{l} \text{(b)} \quad 4x_1 + x_2 - 5x_3 = 8 \\ \quad -2x_1 + 3x_2 + x_3 = 12 \\ \quad 3x_1 - x_2 + 4x_3 = 5 \end{array}$$

7. É possível que uma matriz seja sua própria inversa?

5.5 Regra de Cramer

O método de inversão de matriz discutido na Seção 5.4 nos permite derivar um modo prático, se bem que nem sempre eficiente, de resolver um sistema de equações lineares, conhecido como *regra de Cramer*.

Dedução da regra

Dado um sistema de equações $Ax = d$, onde A é $n \times n$, a solução pode ser escrita como

$$x^* = A^{-1}d = \frac{1}{|A|} (\text{adj } A) d \quad [\text{por (5.16)}]$$

contanto que A seja invertível. De acordo com (5.15), isso significa que

$$\begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \cdots & \cdots & \cdots & \cdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} d_1 |C_{11}| + d_2 |C_{21}| + \cdots + d_n |C_{n1}| \\ d_1 |C_{12}| + d_2 |C_{22}| + \cdots + d_n |C_{n2}| \\ \cdots \\ d_1 |C_{1n}| + d_2 |C_{2n}| + \cdots + d_n |C_{nn}| \end{bmatrix}$$

$$= \frac{1}{|A|} \begin{bmatrix} \sum_{i=1}^n d_i |C_{i1}| \\ \sum_{i=1}^n d_i |C_{i2}| \\ \vdots \\ \sum_{i=1}^n d_i |C_{in}| \end{bmatrix}$$

Igualando os elementos correspondentes dos dois lados da equação, obtemos os valores de solução

$$x_1^* = \frac{1}{|A|} \sum_{i=1}^n d_i |C_{i1}| \quad x_2^* = \frac{1}{|A|} \sum_{i=1}^n d_i |C_{i2}| \quad (\text{etc.}) \quad (5.17)$$

Os termos $\sum$ em (5.17) parecem estranhos. O que significam? De (5.8), vemos que a expansão de Laplace de um determinante $|A|$ por sua primeira coluna pode ser expressa na forma $\sum_{i=1}^n a_{i1} |C_{i1}|$. Se substituirmos a primeira coluna de $|A|$ pelo vetor coluna d , mas mantivermos intactas todas as outras colunas, então resultará um novo determinante, que podemos denominar $|A_1|$ – o índice 1 indica que a primeira coluna foi substituída por d . A expansão de $|A_1|$ por sua primeira coluna (a coluna d) resultará na expressão $\sum_{i=1}^n d_i |C_{i1}|$, porque os elementos d_i agora tomam o lugar dos elementos a_{i1} . Voltando a (5.17), vemos, por conseguinte, que

$$x_1^* = \frac{1}{|A|} |A_1|$$

De maneira semelhante, se substituirmos a segunda coluna de $|A|$ pelo vetor coluna d , mantendo, ao mesmo tempo, todas as outras colunas, a expansão do novo determinante $|A_2|$ por sua segunda coluna (a coluna d) resultará na expressão $\sum_{i=1}^n d_i |C_{i2}|$. Quando dividida por $|A|$, esta última soma nos dará o valor de solução x_2^* e assim por diante.

Esse procedimento agora pode ser generalizado. Para achar o valor de solução da j -ésima variável x_j^* , basta substituir a j -ésima coluna do determinante $|A|$ pelos termos constantes $d_1 \cdots d_n$ para obter um novo determinante $|A_j|$ e então dividir $|A_j|$ pelo determinante original $|A|$. Assim, a solução do sistema $Ax = d$ pode ser expressa como

$$x_j^* = \frac{|A_j|}{|A|} = \frac{1}{|A|} \begin{vmatrix} a_{11} & a_{12} & \cdots & d_1 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & d_2 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & d_n & \cdots & a_{nn} \end{vmatrix} \quad (5.18)$$

(j -ésima coluna substituída por d)

O resultado em (5.18) é o enunciado da regra de Cramer. Note que, ao passo que o método de inversão de matriz resulta nos valores de solução de *todas* as variáveis endógenas de uma só vez (x^* é um vetor), a regra de Cramer pode nos dar o valor de solução de uma única variável endógena por vez (x_j^* é um escalar); é por isso que ela pode não ser eficiente.

EXEMPLO 1 Descubra a solução do sistema de equações

$$5x_1 + 3x_2 = 30$$

$$6x_1 - 2x_2 = 8$$

Os coeficientes e os termos constantes resultam nos seguintes determinantes:

$$|A| = \begin{vmatrix} 5 & 3 \\ 6 & -2 \end{vmatrix} = -28 \quad |A_1| = \begin{vmatrix} 30 & 3 \\ 8 & -2 \end{vmatrix} = -84$$

$$|A_2| = \begin{vmatrix} 5 & 30 \\ 6 & 8 \end{vmatrix} = -140$$

Portanto, em virtude de (5.18), podemos escrever imediatamente

$$x_1^* = \frac{|A_1|}{|A|} = \frac{-84}{-28} = 3 \quad \text{e} \quad x_2^* = \frac{|A_2|}{|A|} = \frac{-140}{-28} = 5$$

EXEMPLO 2 Encontre a solução do sistema de equações

$$7x_1 - x_2 - x_3 = 0$$

$$10x_1 - 2x_2 + x_3 = 8$$

$$6x_1 + 3x_2 - 2x_3 = 7$$

Verificamos que os determinantes relevantes $|A|$ e $|A_j|$ são

$$|A| = \begin{vmatrix} 7 & -1 & -1 \\ 10 & -2 & 1 \\ 6 & 3 & -2 \end{vmatrix} = -61 \quad |A_1| = \begin{vmatrix} 0 & -1 & -1 \\ 8 & -2 & 1 \\ 7 & 3 & -2 \end{vmatrix} = -61$$

$$|A_2| = \begin{vmatrix} 7 & 0 & -1 \\ 10 & 8 & 1 \\ 6 & 7 & -2 \end{vmatrix} = -183 \quad |A_3| = \begin{vmatrix} 7 & -1 & 0 \\ 10 & -2 & 8 \\ 6 & 3 & 7 \end{vmatrix} = -244$$

assim, os valores de solução das variáveis são

$$x_1^* = \frac{|A_1|}{|A|} = \frac{-61}{-61} = 1 \quad x_2^* = \frac{|A_2|}{|A|} = \frac{-183}{-61} = 3 \quad x_3^* = \frac{|A_3|}{|A|} = \frac{-244}{-61} = 4$$

Note que, em cada um desses exemplos, achamos $|A| \neq 0$. Essa é uma condição necessária para a aplicação da regra de Cramer, assim como é uma condição necessária para a existência da inversa A^{-1} . Afinal, a regra de Cramer é baseada no conceito da matriz inversa, mesmo que, na prática, ela evite o processo de inversão de matriz.

Observação sobre sistemas homogêneos de equações

O sistema de equações $Ax = d$ considerado anteriormente pode ter quaisquer constantes no vetor d . Se $d = 0$, isto é, se $d_1 = d_2 = \dots = d_n = 0$, contudo, o sistema de equações se tornará

$$Ax = 0$$

onde 0 é o vetor zero. Esse caso especial é denominado um *sistema homogêneo de equações*. Aqui, a palavra *homogêneo* está relacionada com a propriedade que diz que, quando se multiplicam todas as variáveis $x_1, \dots, x_n$ pelo mesmo número, o sistema de equações continuará válido. Isso é possível somente se os termos constantes do sistema – os termos que não estão ligados a qualquer x_i – forem todos zero.

Se a matriz A for invertível, um sistema homogêneo de equações pode dar somente uma “solução trivial”, a saber, $x_1^* = x_2^* = \dots = x_n^* = 0$. Isso resulta do fato de que, neste caso, a solução $x^* = A^{-1}d$ se tornará

$$x^* = A^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Alternativamente, esse resultado pode ser obtido da regra de Cramer. O fato de $d = 0$ implica que, para todo j , $|A_j|$ deve conter uma coluna inteira de zeros, e assim, a solução será

$$x_j^* = \frac{|A_j|}{|A|} = \frac{0}{|A|} = 0 \quad (j = 1, 2, \dots, n)$$

Muito curioso é que o *único* modo de obter uma solução *não-trivial* de um sistema homogêneo de equações é ter $|A| = 0$, isto é, ter uma matriz de coeficientes *singular*, A ! Nesse caso, temos

$$x_j^* = \frac{|A_j|}{|A|} = \frac{0}{0}$$

onde a expressão $0/0$ não é igual a zero, mas é algo indefinido. Por consequência, a regra de Cramer não é aplicável. Isso não quer dizer que não podemos obter soluções; quer dizer apenas que não podemos obter uma solução única.

Observe o sistema homogêneo de equações

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_1 + a_{22}x_2 &= 0 \end{aligned} \quad (5.19)$$

É evidente, por si só, que $x_1^* = x_2^* = 0$ é uma solução, mas essa solução é trivial. Agora, admita que a matriz de coeficientes A é singular, de modo que $|A| = 0$. Isso implica que o vetor linha $[a_{11} \ a_{12}]$ é um múltiplo do vetor linha $[a_{21} \ a_{22}]$; em consequência, uma das duas equações é redundante. Excluindo, digamos, a segunda equação de (5.19), acabamos com uma equação (a primeira) de duas variáveis, cuja solução é $x_1^* = (-a_{12}/a_{11})x_2^*$. Essa solução é não trivial e bem definida se $a_{11} \neq 0$, mas, na realidade, ela representa um número infinito de soluções porque, para cada valor possível de x_2^* , há um valor correspondente x_1^* tal que o par constitui uma solução. Assim, não existe nenhuma solução não trivial única para esse sistema homogêneo de equações. Esta última afirmação também é, de modo geral, válida para o caso de n variáveis.

Tipos de soluções para um sistema de equações lineares

Nossa discussão sobre as diversas variantes do sistema de equações lineares $Ax = d$ revela que há quatro tipos diferentes de soluções possíveis. Para visualizar melhor essas variantes, elas são apresentadas na Tabela 5.1.

Na primeira possibilidade, o sistema pode ter uma solução única, não trivial. Esse tipo de resultado pode surgir somente quando temos um sistema não homogêneo com uma matriz de coeficientes invertível, A . O segundo resultado possível é uma solução trivial, única, que é associada a um sistema homogêneo com uma matriz A invertível. Na terceira possibilidade, podemos ter um número infinito de soluções. Essa ocorrência está ligada exclusivamente a um sistema no qual as equações são dependentes (isto é, no qual há equações redundantes). Dependendo de o sistema ser ou não homogêneo, a solução trivial pode ou não estar incluída no conjunto infinito de soluções. Por fim, no caso de um sistema de equações inconsistente, não existe absolutamente nenhuma solução. Do ponto de vista de quem monta um modelo, o resultado mais útil e desejável é, obviamente, o de uma solução única, não trivial, $x^* \neq 0$.

TABELA 5.1
Tipos de
soluções para
um sistema de
equações
lineares $Ax = d$

Determinante $ A $	Vetor d	
	$d \neq 0$ (sistema não homogêneo)	$d = 0$ (sistema homogêneo)
$ A \neq 0$ (matriz A invertível)	Existe uma solução trivial única, $x^* \neq 0$.	Existe uma solução trivial única $x^* = 0$.
$ A = 0$ (matriz A singular)		
Equações dependentes	Existe um número infinito de soluções (não incluindo a solução trivial).	Existe um número infinito de soluções (incluindo a solução trivial).
Equações inconsistentes	Não existe solução.	[Não é possível.]

EXERCÍCIO 5.5

- Use a regra de Cramer para resolver os seguintes sistemas de equações:

(a) $3x_1 - 2x_2 = 6$ $2x_1 + x_2 = 11$	(c) $8x_1 - 7x_2 = 9$ $x_1 + x_2 = 3$
(b) $-x_1 + 3x_2 = -3$ $4x_1 - x_2 = 12$	(d) $5x_1 + 9x_2 = 14$ $7x_1 - 3x_2 = 4$
- Para cada um dos sistemas de equações no Problema 1, encontre a inversa da matriz de coeficientes e obtenha a solução pela fórmula $x^* = A^{-1}d$.
- Use a regra de Cramer para resolver os seguintes sistemas de equações:

(a) $8x_1 - x_2 = 16$ $2x_2 + 5x_3 = 5$ $2x_1 + 3x_3 = 7$	(c) $4x + 3y - 2z = 1$ $x + 2y = 6$ $3x + z = 4$
(b) $-x_1 + 3x_2 + 2x_3 = 24$ $x_1 + x_3 = 6$ $5x_2 - x_3 = 8$	(d) $-x + y + z = a$ $x - y + z = b$ $x + y - z = c$
- Mostre que a regra de Cramer pode ser obtida, alternativamente, pelo seguinte procedimento. Multiplique ambos os lados da primeira equação no sistema $Ax = d$ pelo co-fator $|C_1 j|$ e, então, multiplique ambos os lados da segunda equação pelo co-fator $|C_2 j|$ etc. Some todas as equações que obteve. Então, atribua os valores $1, 2, \dots, n$ ao índice j , sucessivamente, para obter os valores de solução $x_1^*, x_2^*, \dots, x_n^*$, como mostrado em (5.17).

5.6 Aplicação aos modelos de mercado e de renda nacional

Modelos simples de equilíbrio como os discutidos no Capítulo 3 podem ser resolvidos com facilidade pela regra de Cramer ou por inversão de matriz.

Modelo de mercado

O modelo de mercado de duas mercadorias descrito em (3.12) pode ser escrito (após a eliminação das variáveis de quantidade) como um sistema de duas equações lineares, como em (3.13'):

$$c_1 P_1 + c_2 P_2 = -c_0$$

$$\gamma_1 P_1 + \gamma_2 P_2 = -\gamma_0$$

Os três determinantes necessários – $|A|$, $|A_1|$ e $|A_2|$ – têm os seguintes valores:

$$|A| = \begin{vmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{vmatrix} = c_1 \gamma_2 - c_2 \gamma_1$$

$$|A_1| = \begin{vmatrix} -c_0 & c_2 \\ -\gamma_0 & \gamma_2 \end{vmatrix} = -c_0 \gamma_2 + c_2 \gamma_0$$

$$|A_2| = \begin{vmatrix} c_1 & -c_0 \\ \gamma_1 & -\gamma_0 \end{vmatrix} = -c_1 \gamma_0 + c_0 \gamma_1$$

Portanto, os preços de equilíbrio devem ser

$$P_1^* = \frac{|A_1|}{|A|} = \frac{c_2 \gamma_0 - c_0 \gamma_2}{c_1 \gamma_2 - c_2 \gamma_1} \quad P_2^* = \frac{|A_2|}{|A|} = \frac{c_0 \gamma_1 - c_1 \gamma_0}{c_1 \gamma_2 - c_2 \gamma_1}$$

que são precisamente os obtidos em (3.14) e (3.15). As quantidades de equilíbrio podem ser encontradas, como antes, estabelecendo $P_1 = P_1^*$ e $P_2 = P_2^*$ nas funções de demanda ou de oferta.

Modelo de renda nacional

O modelo de renda nacional simples citado em (3.23) também pode ser resolvido pela utilização da regra de Cramer. Como escrito em (3.23), o modelo consiste nas duas equações simultâneas seguintes:

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= a + bY \quad (a > 0, \quad 0 < b < 1) \end{aligned}$$

Essas equações podem ser rearranjadas sob a forma

$$\begin{aligned} Y - C &= I_0 + G_0 \\ -bY + C &= a \end{aligned}$$

de modo que as variáveis endógenas Y e C apareçam somente à esquerda dos sinais de igual, enquanto as variáveis exógenas e o parâmetro livre apareçam somente à direita. A matriz de coeficientes agora toma a forma $\begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$ e o vetor coluna de constantes (dados), a forma $\begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix}$. Note que a soma $I_0 + G_0$ é considerada como uma única entidade, isto é, um único elemento no vetor de constantes.

A regra de Cramer agora leva imediatamente à seguinte solução:

$$\begin{aligned} Y^* &= \frac{\begin{vmatrix} (I_0 + G_0) & -1 \\ a & 1 \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{I_0 + G_0 + a}{1 - b} \\ C^* &= \frac{\begin{vmatrix} 1 & (I_0 + G_0) \\ -b & a \end{vmatrix}}{\begin{vmatrix} 1 & -1 \\ -b & 1 \end{vmatrix}} = \frac{a + b(I_0 + G_0)}{1 - b} \end{aligned}$$

Você deve verificar que os valores de solução que acabamos de obter são idênticos aos mostrados em (3.24) e (3.25).

Agora vamos tentar resolver esse modelo invertendo a matriz de coeficientes. Visto que a matriz de coeficientes é $A = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}$, sua matriz de co-fatores é $\begin{bmatrix} 1 & b \\ 1 & 1 \end{bmatrix}$ e, portanto, temos $\text{adj } A =$

$\begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$. Segue-se que a matriz inversa é

$$A^{-1} = \frac{1}{|A|} \text{adj } A = \frac{1}{1 - b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix}$$

Sabemos que, para o sistema de equações $Ax = d$, a solução pode ser expressa como $x^* = A^{-1}d$. Aplicada ao presente modelo, isso significa que

$$\begin{bmatrix} Y^* \\ C^* \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} 1 & 1 \\ b & 1 \end{bmatrix} \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} = \frac{1}{1-b} \begin{bmatrix} I_0 + G_0 + a \\ b(I_0 + G_0) + a \end{bmatrix}$$

É fácil ver que essa é, mais uma vez, a mesma solução obtida anteriormente.

O modelo IS-LM: economia fechada

Para um outro modelo linear da economia, podemos imaginar que a economia seja composta de dois setores: o setor de bens reais e o setor monetário.

O mercado de bens envolve as seguintes equações:

$$Y = C + I + G$$

$$C = a + b(1-t)Y$$

$$I = d - ei$$

$$G = G_0$$

As variáveis endógenas são Y , C , I e i (onde i é a taxa de juros). A variável exógena é G_0 , enquanto a , d , e , b e t são parâmetros estruturais.

No recém-introduzido mercado monetário, temos:

$$\text{Condição de equilíbrio: } M_d = M_s$$

$$\text{Demanda por moeda: } M_d = kY - li$$

$$\text{Oferta de moeda: } M_s = M_0$$

onde M_0 é o estoque exógeno de moeda e k e l são parâmetros. Essas três equações podem ser condensadas em:

$$M_0 = kY - li$$

Juntos, os dois setores nos dão o seguinte sistema de equações:

$$Y - C - I = G_0$$

$$b(1-t)Y - C = -a$$

$$I + ei = d$$

$$kY - li = M_0$$

Note que, fazendo mais substituições, o sistema poderia ser reduzido ainda mais até um sistema de equações 2×2 . Por enquanto, vamos continuar com um sistema 4×4 . Sob a forma de matriz, temos

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ b(1-t) & -1 & 0 & 0 \\ 0 & 0 & 1 & e \\ k & 0 & 0 & -l \end{bmatrix} \begin{bmatrix} Y \\ C \\ I \\ i \end{bmatrix} = \begin{bmatrix} G_0 \\ -a \\ d \\ M_0 \end{bmatrix}$$

Para encontrar o determinante da matriz de coeficientes, podemos usar a expansão de Laplace em uma das colunas (de preferência, a que tiver mais zeros). Expandindo a quarta coluna, obtemos

$$|A| = (-e) \begin{vmatrix} 1 & -1 & -1 \\ b(1-t) & -1 & 0 \\ k & 0 & 0 \end{vmatrix} - l \begin{vmatrix} 1 & -1 & -1 \\ b(1-t) & -1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$\begin{aligned}
&= (-e)(k) \begin{vmatrix} -1 & -1 \\ -1 & 0 \end{vmatrix} - l \begin{vmatrix} 1 & -1 \\ b(1-t) & -1 \end{vmatrix} \\
&= ek - l[(-1) - (-1)b(1-t)] \\
&= ek + l[1 - b(1-t)]
\end{aligned}$$

Podemos usar a regra de Cramer para encontrar a renda de equilíbrio Y^* . Fazemos isso substituindo a primeira coluna da matriz de coeficientes A pelo vetor de variáveis exógenas e tomando a razão entre o determinante da nova matriz e o determinante original, ou

$$Y^* = \frac{|A_1|}{|A|} = \frac{\begin{vmatrix} G_0 & -1 & -1 & 0 \\ -a & -1 & 0 & 0 \\ d & 0 & 1 & e \\ M_0 & 0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]}$$

Usar a expansão de Laplace na segunda coluna do numerador resulta em

$$\begin{aligned}
Y^* &= \frac{(-1)(-1)^3 \begin{vmatrix} -a & 0 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix} + (-1)(-1)^4 \begin{vmatrix} G_0 & -1 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]} \\
&= \frac{\begin{vmatrix} -a & 0 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix} - \begin{vmatrix} G_0 & -1 & 0 \\ d & 1 & e \\ M_0 & 0 & -l \end{vmatrix}}{ek + l[1 - b(1-t)]}
\end{aligned}$$

Expandindo ainda mais, obtemos

$$\begin{aligned}
Y^* &= \frac{(1) \begin{vmatrix} -a & 0 \\ M_0 & -l \end{vmatrix} - \left\{ (-1)(-1)^3 \begin{vmatrix} d & e \\ M_0 & -l \end{vmatrix} + (-1)^4 \begin{vmatrix} G_0 & 0 \\ M_0 & -l \end{vmatrix} \right\}}{ek + l[1 - b(1-t)]} \\
&= \frac{al - [d(-l) - eM_0] - G_0(-l)}{ek + l[1 - b(1-t)]} \\
&= \frac{l(a + d + G_0) + eM_0}{ek + l[1 - b(1-t)]}
\end{aligned}$$

Visto que a solução para Y^* é linear no que diz respeito às variáveis exógenas, podemos reescrever Y^* como

$$Y^* = \left(\frac{e}{ek + l[1 - b(1-t)]} \right) M_0 + \left(\frac{l}{ek + l[1 - b(1-t)]} \right) (a + d + G_0)$$

Sob essa forma, podemos ver que os multiplicadores da política keynesiana relativos à oferta de moeda e às despesas do governo são os coeficientes de M_0 e G_0 , isto é,

Multiplicador da oferta de moeda:

$$\frac{e}{ek + [1 - b(1 - t)]}$$

e

Multiplicador das despesas do governo:

$$\frac{l}{ek + [1 - b(1 - t)]}$$

Álgebra matricial versus eliminação de variáveis

Os modelos econômicos usados para a ilustração anterior envolvem somente duas das quatro equações e, assim, precisamos examinar apenas determinantes de quarta ordem ou de ordem mais baixa. Para grandes sistemas de equações surgirão determinantes de ordens mais altas, e seu cálculo será mais complicado. E mais complicada será a inversão de grandes matrizes. Do ponto de vista do cálculo, na verdade, a inversão de matriz e a regra de Cramer não são necessariamente mais eficientes que o método de eliminação sucessiva de variáveis.

Contudo, métodos matriciais têm outros méritos. Como vimos nas páginas precedentes, a álgebra matricial nos dá uma notação compacta para qualquer sistema de equações lineares e também fornece um critério por meio de determinantes, para testar a existência de uma solução única. Essas vantagens não são conseguidas de outro modo. Além disso, deve-se observar que, diferentemente do método de eliminação de variáveis, que não permite nenhum meio de expressar a solução de modo analítico, o método da inversão de matriz e a regra de Cramer realmente nos fornecem expressões muito úteis para a solução, $x^* = A^{-1}d$ e $x_j^* = |A_j|/|A|$. Essas expressões analíticas da solução são úteis não apenas porque são, em si mesmas, um enunciado resumido do procedimento de solução propriamente dito, mas também porque possibilitam executar outras operações matemáticas com a solução tal como escrita, se for necessário.

Sob certas circunstâncias, podemos até declarar que os métodos matriciais têm uma vantagem de cálculo, por exemplo, quando temos de resolver, ao mesmo tempo, vários sistemas de equações que têm uma matriz de coeficientes, A , idêntica, mas vetores de termos constantes diferentes. Nesse caso, o método de eliminação de variáveis exigiria que o procedimento de cálculo fosse repetido toda vez que um novo sistema de equações fosse considerado. Com o método de inversão de matrizes, entretanto, temos de achar a matriz inversa comum A^{-1} *somente uma vez*; então, a mesma inversa pode ser usada para pré-multiplicar todos os vetores de termos constantes pertinentes aos vários sistemas de equações envolvidos, de modo a obter suas respectivas soluções. Essa vantagem específica de cálculo adquirirá grande significância prática quando considerarmos a solução dos modelos de insumo-produto de Leontief na Seção 5.7.

EXERCÍCIO 5.6

- Resolva o modelo de renda nacional no Exercício 3.5-1:
 - Por inversão de matriz
 - Pela regra de Cramer
 (Relacione as variáveis na ordem Y, C, T .)
- Resolva o modelo de renda nacional no Exercício 3.5-2:
 - Por inversão de matriz
 - Pela regra de Cramer
 (Relacione as variáveis na ordem Y, C, G .)
- Seja a equação IS

$$Y = \frac{A}{1-b} - \frac{g}{1-b}i$$

onde $1 - b$ é a propensão marginal a poupar, g é a sensibilidade do investimento às taxas de juros e A é um agregado de variáveis exógenas. Seja a equação LM

$$Y = \frac{M_0}{k} + \frac{l}{k}i$$

onde k e l são a renda e a sensibilidade aos juros da demanda de moeda, respectivamente e M_0 são os estoques monetários reais.

Se $b = 0,7$, $g = 100$, $A = 252$, $k = 0,25$, $l = 200$ e $M_0 = 176$, então

- (a) Escreva o sistema IS-LM sob a forma de matriz.
(b) Calcule Y e i por inversão de matriz.

5.7 Modelos de insumo-produto de Leontief

Em sua versão “estática”, a análise de insumo-produto do Professor Wassily Leontief, ganhador do Prêmio Nobel,[†] trata da seguinte pergunta, em particular: “Que nível de produto cada uma das n indústrias de uma economia deve produzir, de modo que seja exatamente suficiente para satisfazer a demanda total por aquele produto?”

O princípio racional para o termo *análise de insumo-produto* é muito fácil de perceber. O produto de qualquer indústria (digamos, a indústria siderúrgica) é necessário como insumo para muitas outras indústrias, ou mesmo para a própria indústria; portanto, o nível “correto” (isto é, sem escassez nem excesso) de produção de aço dependerá dos requisitos de insumos de todas as n indústrias. Por sua vez, os produtos de muitas outras indústrias entrarão na indústria siderúrgica como insumos e, por consequência, os níveis “corretos” dos outros produtos, por sua vez, dependerão parcialmente dos requisitos de insumos da indústria siderúrgica. Em vista dessa dependência entre indústrias, qualquer conjunto de níveis “corretos” de produtos para as n indústrias deve ser consistente com todos os requisitos de insumos na economia, de modo que nenhum gargalo surgirá em nenhum lugar. Sob essa perspectiva, é claro que a análise de insumo-produto deve ser de grande utilidade para o planejamento da produção, tal como o planejamento do desenvolvimento econômico de um país ou de um programa de defesa nacional.

A rigor, a análise de insumo-produto não é uma forma da análise do equilíbrio geral como discutida no Capítulo 3. Embora a interdependência das várias indústrias seja enfatizada, os níveis “corretos” de produtos contemplados são aqueles que satisfazem relações técnicas insumo-produto, e não condições de equilíbrio de mercado. Não obstante, o problema proposto pela análise de insumo-produto também se reduz a um problema de resolver um sistema de equações simultâneas e, mais uma vez, a álgebra matricial pode ser útil.

Estrutura de um modelo de insumo-produto

Visto que um modelo de insumo-produto normalmente abrange um grande número de indústrias, sua estrutura é, necessariamente, bastante complicada. Para simplificar o problema, as seguintes premissas são adotadas em geral: (1) cada indústria produz somente uma mercadoria homogênea (por uma interpretação ampla, isso permite o caso de duas ou mais mercadorias produzidas em conjunto, contanto que sejam produzidas segundo uma proporção fixa entre uma e outra); (2) cada indústria utiliza uma razão fixa de insumos (ou combinação de fatores) para a produção de seu produto; e (3) a produção de cada uma das indústrias está sujeita a retornos constantes de escala, de modo que uma mudança de k vezes em cada insumo resultará em uma mudança de exatamente k vezes no produto. É claro que essas premissas não são realistas. Mas a salvação é que, se uma indústria produz duas mercadorias diferentes ou utiliza duas possíveis combinações de fatores, então essa indústria pode – ao menos conceitualmente – ser dividida em duas indústrias separadas.

Vemos, por essas premissas que, para produzir cada unidade da j -ésima mercadoria, a quantidade de insumo para a i -ésima mercadoria tem de ser fixa, o que denotaremos por a_{ij} . Especificamente, a produção de cada unidade da j -ésima mercadoria requerirá a_{1j} (quantidade) da primeira mercadoria, a_{2j} da segunda mercadoria, $\dots$ e a_{nj} da n -ésima mercadoria (A ordem dos índices em a_{ij} é fácil de lembrar: o primeiro índice se refere ao insumo e o segundo ao produto, de modo que a_{ij} indica quanto da i -ésima mercadoria é utilizado na produção de cada unidade da

[†] Leontief, Wassily W. *The Structure of American Economy 1919–1939*. 2. ed., Fair Lawn, N.J.: Oxford University Press, 1951.

j -ésima mercadoria). Para nossa finalidade, podemos admitir que os preços são dados e, assim, adotar um “valor em dólar” de cada mercadoria como sua unidade. Então, a declaração $a_{32} = 0,35$ significará que é requerido um valor de 35 centavos da terceira mercadoria como insumo para produzir o valor de um dólar da segunda mercadoria. O símbolo a_{ij} será denominado um *coeficiente de insumo*.

Para uma economia de n indústrias, os coeficientes de insumos podem ser organizados em uma matriz $A = [a_{ij}]$, como apresentado na Tabela 5.2, na qual cada *coluna* especifica os requisitos de insumos para a produção de uma unidade do produto de uma indústria em particular. A segunda coluna, por exemplo, declara que, para produzir uma unidade (o valor de um dólar) da mercadoria II, os insumos necessários são: a_{12} unidades da mercadoria I, a_{22} unidades da mercadoria II etc. Se nenhuma indústria utilizar seu próprio produto como insumo, então os elementos na diagonal principal da matriz A serão todos zero.

O modelo aberto

Se as n indústrias da Tabela 5.2 constituíssem a totalidade da economia, então todos os seus produtos teriam o único propósito de atender à *demanda de insumos* das mesmas n indústrias (ser utilizados em mais produção), e não à *demanda final* (tal como a demanda de consumo, que não é mais produção). Ao mesmo tempo, todos os insumos utilizados na economia teriam a natureza de *insumos intermediários* (os fornecidos pelas n indústrias), e não de *insumos primários* (tal como mão-de-obra, que não é um produto industrial). Para levar em conta a presença de demanda final e insumos primários, devemos incluir no modelo um *setor aberto* externo à rede de n indústrias. Esse setor aberto pode acomodar as atividades dos domicílios de consumidores, do setor governamental e até de países estrangeiros.

Em vista da presença do setor aberto, a soma dos elementos em cada coluna da matriz de coeficientes de insumos A (*matriz de insumos* A , de forma abreviada) deve ser menor que 1. A soma de cada coluna representa o custo *parcial* do insumo (sem incluir o custo dos insumos primários) incorrido na produção de um dólar de uma mercadoria; se essa soma for maior ou igual a \$1, portanto, a produção não será economicamente justificável. Esse fato pode ser enunciado simbolicamente assim:

$$\sum_{i=1}^n a_{ij} < 1 \quad (j = 1, 2, \dots, n)$$

onde o somatório é em i , isto é, nos elementos que aparecem nas várias *linhas* de uma coluna específica j . Prosseguindo nessa mesma linha de pensamento, pode-se afirmar, também, que, uma vez que o valor do produto (\$1) deve ser totalmente absorvido pelos pagamentos para todos os fatores de produção, a quantidade que faltar entre a soma da coluna e \$1 deve representar o pagamento dos insumos primários do setor aberto. Assim, o valor dos insumos primários necessários para produzir uma unidade da j -ésima mercadoria seria $1 - \sum_{i=1}^n a_{ij}$.

Insumo	Produto				
	I	II	III	...	N
I	a_{11}	a_{12}	a_{13}	...	a_{1n}
II	a_{21}	a_{22}	a_{23}	...	a_{2n}
III	a_{31}	a_{32}	a_{33}	...	a_{3n}
...	...	...	...	...	...
N	a_{n1}	a_{n2}	a_{n3}	...	a_{nn}

TABELA 5.2
Matriz de
coeficientes
de insumo

Se quisermos que a indústria I fabrique exatamente suficiente para atender aos requisitos de insumos das n indústrias, bem como à demanda final do setor aberto, seu nível de produto x_1 deve satisfazer à seguinte equação:

$$x_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + d_1$$

onde d_1 denota a demanda final para seu produto e $a_{1j}x_j$ representa a demanda de insumos da j -ésima indústria.[†] Pelo mesmo critério, os níveis de produto das outras indústrias devem satisfazer as equações

$$x_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + d_2$$

$$\dots\dots\dots$$

$$x_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + d_n$$

Após transferir todos os termos que envolvem as variáveis x_j para a esquerda do sinal de igual e deixar à direita somente as demandas finais determinadas exogenamente, d_j , podemos expressar os níveis “corretos” de produção das n indústrias pelo seguinte sistema de n equações lineares:

$$\begin{aligned} (1 - a_{11})x_1 - a_{12}x_2 - \cdots - a_{1n}x_n &= d_1 \\ -a_{21}x_1 + (1 - a_{22})x_2 - \cdots - a_{2n}x_n &= d_2 \\ \dots\dots\dots \\ -a_{n1}x_1 - a_{n2}x_2 - \cdots + (1 - a_{nn})x_n &= d_n \end{aligned} \quad (5.20)$$

que pode ser escrito em notação matricial como

$$\begin{bmatrix} (1 - a_{11}) & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & (1 - a_{22}) & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & (1 - a_{nn}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \quad (5.20')$$

Se os 1s na diagonal principal da matriz à esquerda forem ignorados, a matriz é simplesmente $-A = [-a_{ij}]$. Por outro lado, na forma em que está, a matriz é a soma da matriz identidade I_n (com 1s em sua diagonal principal e 0s em todos os outros lugares) e da matriz $-A$. Assim, (5.20') também pode ser escrita como

$$(I - A)x = d \quad (5.20'')$$

onde x e d são, respectivamente, o vetor de variáveis e o vetor de demanda final (termo constante). A matriz $I - A$ é denominada *matriz de Leontief*. Contanto que $I - A$ seja invertível, poderemos achar sua inversa $(I - A)^{-1}$ e obter a solução única do sistema a partir da equação

$$x^* = (I - A)^{-1}d \quad (5.21)$$

Um exemplo numérico

Para fins de ilustração, suponha que há somente três indústrias na economia e um único insumo primário, e que a matriz de coeficientes de insumos é a seguinte (desta vez, usaremos valores decimais):

[†] Nunca adicione os coeficientes de insumos de uma linha. Tal soma – por exemplo, $a_{11} + a_{12} + \cdots + a_{1n}$ – não tem nenhum significado econômico útil. A soma dos produtos $a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$, por outro lado, tem um significado econômico; representa a quantidade total de x_1 que todas as n indústrias necessitam como insumo.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0,2 & 0,3 & 0,2 \\ 0,4 & 0,1 & 0,2 \\ 0,1 & 0,3 & 0,2 \end{bmatrix} \quad (5.22)$$

Note que a soma de cada coluna em A é menor que 1, como deve ser. Além disso, se denotarmos por a_{0j} a quantidade em dólares do insumo primário utilizado para produzir o valor de um dólar da j -ésima mercadoria, podemos escrever [subtraindo de 1 a soma de cada coluna em (5.22)]:

$$a_{01} = 0,3 \quad a_{02} = 0,3 \quad \text{e} \quad a_{03} = 0,4 \quad (5.23)$$

Com a matriz A de (5.22), o sistema de insumo-produto aberto pode ser expresso na forma $(I - A)x = d$ como se segue:

$$\begin{bmatrix} 0,8 & -0,3 & -0,2 \\ -0,4 & 0,9 & -0,2 \\ -0,1 & -0,3 & 0,8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (5.24)$$

Aqui, deliberadamente não atribuímos valores específicos para as demandas finais d_1 , d_2 , e d_3 . Desse modo, mantendo o vetor d em forma paramétrica, nossa solução aparecerá como uma “fórmula” na qual podemos inserir vários vetores d específicos para obter soluções específicas correspondentes.

Invertendo a matriz de Leontief 3×3 , pode-se achar a solução aproximada (devido ao arredondamento dos números decimais) de (5.24):

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = (I - A)^{-1}d = \frac{1}{0,384} \begin{bmatrix} 0,66 & 0,30 & 0,24 \\ 0,34 & 0,62 & 0,24 \\ 0,21 & 0,27 & 0,60 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Se o vetor específico da demanda final (digamos, a meta de produção final de um programa de desenvolvimento) por acaso for $d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$, em bilhões de dólares, teremos (também em bilhões de dólares), os seguintes valores de solução específicos:

$$x_1^* = \frac{1}{0,384} [0,66(10) + 0,30(5) + 0,24(6)] = \frac{9,54}{0,384} = 24,84$$

e, de maneira semelhante,

$$x_2^* = \frac{7,94}{0,384} = 20,68 \quad \text{e} \quad x_3^* = \frac{7,05}{0,384} = 18,36$$

Agora surge uma pergunta importante. A produção da combinação de produtos x_1^* , x_2^* e x_3^* deve acarretar uma quantidade requerida de insumo primário definida. A quantidade *requerida* seria consistente com o que está *disponível* na economia? Com base em (5.23), o insumo primário requerido pode ser calculado como se segue:

$$\sum_{j=1}^3 a_{0j} x_j^* = 0,3(24,84) + 0,3(20,68) + 0,4(18,36) = \$21 \text{ bilhões}$$

Por conseguinte, a demanda final específica $d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}$ será viável se e somente se a quantidade dis-

ponível do insumo primário for, no mínimo, \$21 bilhões. Se a quantidade disponível for menor, é evidente que aquela meta particular de produção terá de ser revista e reduzida de acordo com isso.

Uma característica notável da análise precedente é que, contanto que os coeficientes de insumos permaneçam os mesmos, a inversa, $(I - A)^{-1}$, não mudará; portanto, é preciso executar apenas *uma* inversão de matriz, mesmo que tenhamos de considerar uma centena ou um milhar de vetores de demanda final diferentes – tal como um leque de metas alternativas de desenvolvimento. Isso poupa esforço de cálculo em comparação com o método de eliminação de variáveis. Entretanto, essa vantagem não é compartilhada pela regra de Cramer como esboçada em (5.18) porque, toda vez que for utilizado um vetor de demanda final d diferente, temos de calcular um novo determinante para figurar no numerador em (5.18), o que não é tão simples quanto multiplicar uma matriz inversa conhecida $(I - A)^{-1}$ por um novo vetor d .

A existência de soluções não negativas

No exemplo numérico anterior, a matriz de Leontief $I - A$ é invertível, portanto, existem valores de solução de variáveis de produto, x_j . Além disso, todos os valores de solução x_j^* resultam não negativos, como determinaria o bom-senso econômico. Entretanto, não se pode esperar que esses resultados desejados surjam automaticamente; eles acontecem somente quando a matriz de Leontief possui certas propriedades. Essas propriedades são descritas na denominada *condição de Hawkins-Simon*.[†]

Para explicar essa condição, precisamos introduzir o conceito matemático de *menores principais* de uma matriz, porque os sinais algébricos de menores principais de uma matriz fornecerão pistas importantes que servirão de guia para nossas conclusões analíticas. Já sabemos que, dada uma matriz quadrada, digamos, B , cujo determinante é $|B|$, um menor é um subdeterminante obtido pela eliminação da i -ésima linha e da j -ésima coluna de $|B|$, onde i e j não são necessariamente iguais. Se agora impusermos a restrição de que $i = j$, então o menor resultante é conhecido como o menor *principal*. Por exemplo, dada uma matriz 3×3 B , geralmente podemos escrever seu determinante como

$$|B| = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \quad (5.25)$$

A eliminação simultânea da i -ésima linha e da i -ésima coluna ($i = 3, 2, 1$, sucessivamente) resulta nos seguintes menores principais 2×2 :

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \quad \begin{vmatrix} b_{11} & b_{13} \\ b_{31} & b_{33} \end{vmatrix} \quad \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} \quad (5.26)$$

Em vista de suas dimensões 2×2 , eles são denominados *menores principais de segunda ordem*. Também podemos gerar *menores principais de primeira ordem* (1×1) eliminando de $|B|$ quaisquer duas linhas e as colunas de mesma numeração, que são:

$$|b_{11}| = b_{11} \quad |b_{22}| = b_{22} \quad |b_{33}| = b_{33} \quad (5.27)$$

[†] Hawkins, David e Simon, Herbert A. "Note: Some Conditions of Macroeconomic Stability". *Econometrica*, pp. 245-48. jun./out. de 1949.

Por fim, para completar o quadro, podemos considerar o próprio $|B|$ como o *menor principal de terceira ordem de $|B|$* . Note que, em todos os menores listados em (5.25) até (5.27), os elementos de suas diagonais principais consistem exclusivamente dos elementos da diagonal principal de B . E é aqui que está o princípio racional para o nome “menores principais”[†]

Embora certas aplicações econômicas requeiram a verificação dos sinais algébricos de *todos* os menores principais de uma matriz B , muito freqüentemente nossa conclusão depende somente do padrão de sinais de um subconjunto particular dos menores principais denominados, variadamente, como *menores principais líderes*, *menores principais naturalmente ordenados* ou *menores principais sucessivos*. No caso 3×3 , esse subconjunto consiste somente nos primeiros membros de (5.25) até (5.27):

$$|B_1| \equiv |b_{11}| \quad |B_2| \equiv \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \quad |B_3| \equiv \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \quad (5.28)$$

Aqui, diferente da utilização de índices no contexto da regra de Cramer, é empregado um único índice m no símbolo $|B_m|$ para indicar que o menor principal líder é de dimensão $m \times m$. Um modo fácil de derivar os menores principais líderes é subdividir o determinante $|B|$ com várias linhas tracejadas como mostrado a seguir:

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \quad (5.29)$$

Tomando apenas o elemento de cima na diagonal principal de $|B|$ por si só, temos $|B_1|$; tomando os primeiros dois elementos na diagonal principal, b_{11} e b_{22} , juntamente com os elementos de fora da diagonal que os acompanham, temos $|B_2|$; e assim por diante.

Dado um determinante de ordem mais alta, digamos, $n \times n$, é claro que haverá um número maior de menores principais, mas o seu padrão de construção será o mesmo. Um menor principal de ordem k é sempre obtido pela eliminação de quaisquer $n - k$ linhas e de colunas com a mesma numeração de $|B|$. E seus menores principais líderes $|B_m|$ (com $m = 1, 2, \dots, n$) são sempre formados tomando os primeiros m elementos da diagonal principal em $|B|$ juntamente com os elementos de fora da diagonal que os acompanham.

Com esses subsídios, estamos prontos para enunciar o seguinte teorema importante que se deve a Hawkins e Simon:

Dados (a) uma matriz $n \times n$, B , com $b_{ij} \leq 0$ ($i \neq j$) (isto é, na qual todos os elementos de fora da diagonal são não positivos); e (b) um vetor $n \times 1$, $d \geq 0$ (todos os elementos não negativos), existe um vetor $n \times 1$, $x^* \geq 0$, tal que $Bx^* = d$, se e somente se

[†] Uma definição alternativa de menores principais levaria em conta as várias permutações dos índices i, j e k . No contexto de insumo-produto, isso significaria renumerar as indústrias (por exemplo, a primeira indústria se torna a segunda, e vice-versa, de modo que o índice 11 se torne 22, e o índice 22 se torne 11, e assim por diante). O resultado é que, além dos menores principais 2×2 em (5.26), também teríamos

$$\begin{vmatrix} b_{22} & b_{21} \\ b_{12} & b_{11} \end{vmatrix} \quad \begin{vmatrix} b_{33} & b_{31} \\ b_{13} & b_{11} \end{vmatrix} \quad \text{e} \quad \begin{vmatrix} b_{33} & b_{32} \\ b_{23} & b_{22} \end{vmatrix}$$

Mas esses três últimos, na ordem dada, são exatamente iguais aos três listados em (5.26) em valor e em sinal algébrico; assim, eles podem deixar de ser considerados para nossas finalidades. De modo semelhante, mesmo que a permutação de índices possa gerar menores principais 3×3 adicionais, eles meramente duplicam o menor em (5.25) em valor e em sinal, portanto, também podem ser desconsiderados.

$$|B_m| > 0 \quad (m = 1, 2, \dots, n)$$

isto é, se e somente se os menores principais líderes de B forem todos positivos.

A relevância desse teorema para a análise insumo-produto fica clara quando deixamos B representar a matriz de Leontief $I - A$ (onde $b_{ij} = -a_{ij}$ para $i \neq j$ são, de fato, todos não positivos) e d , o vetor de demanda final (onde todos os elementos são, de fato, não negativos). Então, $Bx^* = d$ é equivalente a $(I - A)x^* = d$, e a existência de $x^* \geq 0$ garante soluções não negativas para os níveis de produção. A condição necessária e suficiente para isso, conhecida como a *condição de Hawkins-Simon* é que todos os menores principais da matriz de Leontief $I - A$ sejam positivos.

A prova desse teorema é muito longa para ser apresentada aqui,[†] mas vale a pena explorar seu significado econômico, que é relativamente fácil de ver no caso simples de duas indústrias ($n = 2$).

Significado econômico da condição de Hawkins-Simon

Para o caso de duas indústrias, a matriz de Leontief é

$$I - A = \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix}$$

A primeira parte da condição de Hawkins-Simon, $|B_1| > 0$, requer que

$$1 - a_{11} > 0 \quad \text{ou} \quad a_{11} < 1$$

Em termos econômicos, isso requer que a quantidade da primeira mercadoria utilizada na produção de um dólar da primeira mercadoria seja menos que um dólar. A outra parte da condição, $|B_2| > 0$, requer que

$$(1 - a_{11})(1 - a_{22}) - a_{12}a_{21} > 0$$

ou o equivalente

$$a_{11} + a_{12}a_{21} + (1 - a_{11})a_{22} < 1$$

Além disso, visto que $(1 - a_{11})a_{22}$ é positivo, a desigualdade precedente implica que

$$a_{11} + a_{12}a_{21} < 1$$

Em termos econômicos, a_{11} mede a utilização *direta* da primeira mercadoria como insumo na produção da própria primeira mercadoria, e $a_{12}a_{21}$ mede a utilização *indireta* – dá a quantidade da primeira mercadoria necessária para produzir a quantidade específica da segunda mercadoria que entra na produção de um dólar da primeira mercadoria. Portanto, a última desigualdade impõe que a quantidade da primeira mercadoria utilizada como insumos direto e indireto na produção de um dólar da mercadoria em si tem de ser menos que um dólar. Assim, o que a condição de Hawkins-Simon faz é especificar certas restrições de praticidade e viabilidade para o processo de produção. O processo de produção pode resultar em soluções não negativas significativas para os níveis de produção, se e somente se for economicamente viável.

[†] Uma discussão completa pode ser encontrada em Takayama, Akira. *Mathematical Economics*. 2. ed., Cambridge University Press, p. 380–385, 1985.

Alguns autores usam uma versão alternativa da condição de Hawkins-Simon, que requer que *todos* os menores principais de B (não somente os líderes) sejam positivos. Entretanto, como Takayama mostra, no caso presente, com a restrição especial imposta a $|B|$, exigir a positividade dos menores principais líderes (uma condição menos restritiva) pode alcançar o mesmo resultado. Não obstante, deve-se enfatizar que, como regra geral, o fato de os menores principais líderes satisfazerem uma determinada exigência de sinal não garante que todos os menores principais automaticamente também satisfaçam aquela exigência. Por consequência, uma condição enunciada em termos de *todos* os menores principais deve ser verificada em relação a *todos* os menores principais, e não apenas aos líderes.

O modelo fechado

Se o setor exógeno do modelo aberto de insumo-produto for absorvido no sistema simplesmente como uma outra *indústria*, o modelo se tornará um *modelo fechado*. Neste modelo, a demanda final e o insumo primário não aparecem; em seu lugar estarão os requisitos de insumos e o produto da indústria recém-concebida. Agora, todos os bens serão *intermediários* por natureza, porque tudo que for produzido é produzido apenas para satisfazer os requisitos de insumos das $(n + 1)$ indústrias no modelo.

A primeira vista, a conversão do setor aberto em uma indústria adicional aparentemente não criaria nenhuma mudança significativa na análise. Contudo, na realidade, visto que admitte-se que a nova indústria tem uma razão fixa de insumos, como teria qualquer outra indústria, o suprimento daquilo que costumava ser insumo primário agora tem de ter uma proporção fixa em relação ao que costumava ser denominado *demanda final*. De modo mais concreto, isso pode significar, por exemplo, que os domicílios consumirão cada mercadoria segundo uma proporção fixa em relação à mão-de-obra que fornecem. Isso constitui, sem dúvida, uma mudança significativa na estrutura analítica envolvida.

Matematicamente, o desaparecimento das demandas finais significa que agora haverá um sistema homogêneo de equações. Admitindo-se somente quatro indústrias (incluindo a nova, designada pelo índice 0), os níveis “corretos” de produção serão, por analogia com (5.20’), aqueles que satisfarão o sistema de equações:

$$\begin{bmatrix} (1 - a_{00}) & -a_{01} & -a_{02} & -a_{03} \\ -a_{10} & (1 - a_{11}) & -a_{12} & -a_{13} \\ -a_{20} & -a_{21} & (1 - a_{22}) & -a_{23} \\ -a_{30} & -a_{31} & -a_{32} & (1 - a_{33}) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Como esse sistema de equações é homogêneo, ele pode ter uma solução não trivial se e somente se a matriz 4×4 de Leontief, $I - A$, tiver um determinante nulo. Esta última condição é, de fato, sempre satisfeita: em um modelo fechado, não existe nenhum insumo primário; daí, cada soma de coluna na matriz de coeficientes de insumos A agora deve ser exatamente igual a 1 (em vez de ser menor que 1); isto é, $a_{0j} + a_{1j} + a_{2j} + a_{3j} = 1$ ou

$$a_{0j} = 1 - a_{1j} - a_{2j} - a_{3j}$$

Mas isso implica que, em cada coluna da matriz $I - A$, dada anteriormente, o elemento de cima é sempre igual ao negativo da soma dos outros três elementos. Por conseqüência, as quatro linhas são linearmente dependentes e temos $|I - A| = 0$. Isso garante que o sistema realmente possui soluções não triviais; de fato, como indicado na Tabela 5.1, ele tem um número infinito delas, o que significa que, em um modelo fechado, com um sistema homogêneo de equações, não existe uma única combinação de produção “correta”. Podemos determinar os níveis de produção $x_1^*, \dots, x_4^*$ proporcionalmente um ao outro, mas não podemos fixar seus níveis absolutos a menos que sejam impostas restrições adicionais ao modelo.

EXERCÍCIO 5.7

1. Com base no modelo em (5.24), se as demandas finais forem $d_1 = 30$, $d_2 = 15$ e $d_3 = 10$ (tudo em bilhões de dólares), quais são as soluções de níveis de produção para as três indústrias? (Arredonde as respostas até duas casas decimais).
2. Usando a informação em (5.23), calcule a quantidade total de insumo primário requerida para produzir os níveis de produção da solução do Problema 1.
3. Em uma economia de duas indústrias, sabe-se que a indústria I usa 10 centavos de seu próprio produto e 60 centavos da mercadoria II para produzir o equivalente a um dólar da mercadoria I; a indústria II não usa nenhum de seus próprios produtos, mas usa 50 centavos da mercadoria I para produzir o equivalente a um dólar da mercadoria II; e o setor aberto demanda \$1 bilhão da mercadoria I e \$2 bilhão da mercadoria II.

- (a) Escreva a matriz de insumos, a matriz de Leontief e a equação matricial insumo-produto para essa economia.
 (b) Verifique se os dados deste problema satisfazem a condição de Hawkins-Simon.
 (c) Encontre os níveis de produção de solução pela regra de Cramer.
4. Dados a matriz de insumos e o vetor de demanda final

$$A = \begin{bmatrix} 0,05 & 0,25 & 0,34 \\ 0,33 & 0,10 & 0,12 \\ 0,19 & 0,38 & 0 \end{bmatrix} \quad d = \begin{bmatrix} 1800 \\ 200 \\ 900 \end{bmatrix}$$

- (a) Explique o significado econômico dos elementos 0,33, 0 e 200.
 (b) Explique o significado econômico (se houver) da soma da terceira coluna.
 (c) Explique o significado econômico (se houver) da soma da terceira linha.
 (d) Escreva a equação matricial de insumo-produto específica para esse modelo.
 (e) Verifique se os dados deste problema satisfazem a condição de Hawkins-Simon.
5. (a) Dada uma matriz 4×4 $B = [b_{ij}]$, escreva todos os seus menores principais.
 (b) Escreva todos os seus menores principais líderes.
6. Mostre que, por si só (sem outras restrições impostas à matriz B), a condição de Hawkins-Simon já garante a existência de um único vetor solução x^* , embora não seja necessariamente não negativo.

5.8 Limitações da análise estática

Na discussão do equilíbrio estático de mercado ou da renda nacional, nossa preocupação primordial tem sido achar os valores de equilíbrio das variáveis endógenas no modelo. Um ponto fundamental que foi ignorado nessa análise é o processo propriamente dito de ajustes e reajustes das variáveis que, por fim, leva ao estado de equilíbrio (se é que ele pode ser alcançado). Nos limitamos apenas a perguntar onde chegaremos, mas não perguntamos quando nem o que pode acontecer no caminho.

A análise do tipo estática, portanto, deixa de levar em conta dois problemas importantes. Um deles, visto que o processo de ajuste pode levar um longo tempo para ser concluído, é que um estado de equilíbrio determinado dentro de uma estrutura particular de análise estática pode ter perdido sua relevância antes mesmo de ser atingido se, no ínterim, forças exógenas ao modelo tiverem sofrido algumas mudanças. Esse é o problema das mudanças do estado de equilíbrio. O segundo é que, mesmo permitindo que o processo de ajuste transcorra sem perturbações, o estado de equilíbrio visualizado em uma análise estática pode ser de todo inalcançável. Esse seria o caso de um estado denominado equilíbrio instável, que é caracterizado pelo fato de que o processo de ajuste levará as variáveis para longe ainda daquele estado de equilíbrio, em vez de trazê-las progressivamente para perto. Por conseguinte, desprezar o processo de ajuste equivale a admitir que não existe o problema da possibilidade de atingir o equilíbrio.

As mudanças do estado de equilíbrio (em resposta a mudanças exógenas) são pertinentes a um tipo de análise denominada *estática comparativa*, e a questão da possibilidade de alcançar o estado de equilíbrio e de sua estabilidade pertence ao domínio da *análise dinâmica*. Cada uma delas serve claramente para preencher uma lacuna significativa na análise estática e, assim, é imperativo examinar também essas áreas de análise. Deixaremos o estudo da análise dinâmica para a Parte 5 do livro e, a seguir, voltaremos nossa atenção para o problema da estática comparativa.

PARTE 3

Análise estática comparativa



CAPÍTULO 6

Estática comparativa e o conceito de derivada

Este capítulo e os Capítulos 7 e 8 serão dedicados aos métodos de análise estática comparativa.

6.1 A natureza da estática comparativa

A estática comparativa, como o nome sugere, diz respeito à comparação de diferentes estados de equilíbrio que são associados a diferentes conjuntos de valores de parâmetros e variáveis exógenas. Para o propósito dessa comparação, sempre começamos admitindo um dado estado inicial de equilíbrio. No modelo do mercado isolado, por exemplo, esse equilíbrio inicial será representado por um preço determinado P^* e por uma quantidade correspondente Q^* . De modo semelhante, no modelo simples de renda nacional de (3.23), o equilíbrio inicial será especificado por um Y^* determinado e um C^* correspondente. Agora, se deixarmos ocorrer uma variação que desequilibre o modelo – sob a forma de uma variação no valor de algum parâmetro ou variável exógena – é claro que o equilíbrio inicial será perturbado. O resultado é que diversas variáveis endógenas terão de passar por certos ajustes. Se admitirmos que pode ser definido um novo estado de equilíbrio relevante aos novos valores dos dados, a pergunta que a análise estática comparativa faz é: como esse novo equilíbrio se compara com o antigo?

Deve-se notar que, na análise comparativa estática, ainda desprezamos o processo de ajuste das variáveis; a comparação será meramente entre o estado de equilíbrio inicial (*antes* da variação) e o estado de equilíbrio final (*após* a variação). Além disso, ainda eliminamos a possibilidade de instabilidade de equilíbrio, pois admitimos que o novo equilíbrio pode ser alcançado, exatamente como fizemos com o antigo.

Um análise estática comparativa pode ser de natureza qualitativa ou quantitativa. Se estivermos interessados apenas na questão, digamos, de saber se um aumento nos investimentos I_0 aumentará ou reduzirá a renda de equilíbrio Y^* , a análise será qualitativa, porque a única questão considerada é a *direção* da variação. Mas, se estivermos preocupados com a *magnitude* (*grandeza*) da variação em Y^* resultante de uma dada variação em I_0 (isto é, com o tamanho do multiplicador do investimento), a análise obviamente será quantitativa. Obtendo uma resposta quantitativa, entretanto, podemos automaticamente determinar a direção da variação pelo seu sinal algébrico. Por conseguinte, a análise quantitativa sempre engloba a qualitativa.

Deve ficar claro que o problema em consideração é, em essência, um problema de encontrar uma *taxa de variação*: a taxa de variação do valor de equilíbrio de uma variável endógena relativa à variação em um determinado parâmetro ou variável exógena. Por essa razão, o conceito matemático de *derivada* assume uma significância preponderante em estática comparativa

porque esse conceito – o mais fundamental do ramo da matemática conhecido como *cálculo diferencial* – diz respeito diretamente à noção de taxa de variação! Além disso, mais adiante veremos que o conceito de derivada também é de extrema importância para problemas de otimização.

6.2 Taxa de variação e a derivada

Mesmo que nosso contexto atual se preocupe somente com as taxas de variação de valores de equilíbrio das variáveis em um modelo, podemos conduzir a discussão de um modo mais geral considerando a taxa de variação de qualquer variável y em resposta a uma variação em uma outra variável x , quando as duas variáveis estão relacionadas uma com a outra pela função

$$y = f(x)$$

Quando aplicada ao contexto da estática comparativa, a variável y representará o valor de equilíbrio de uma variável endógena, e x será um parâmetro. Observe que, de início, estamos nos restringindo ao caso simples em que há somente um único parâmetro ou variável exógena no modelo. Tão logo tenhamos dominado esse caso simplificado, entretanto, a extensão para o caso de mais parâmetros será relativamente fácil.

O quociente de diferenças

Visto que a noção de “variação” figura de modo proeminente no atual contexto, é preciso um símbolo especial para representá-la. Quando a variável x muda de um valor x_0 para um novo valor x_1 , a mudança é medida pela diferença $x_1 - x_0$. Daí, usando o símbolo Δ (a letra grega delta maiúscula, para “diferença”) para denotar a variação, podemos escrever $\Delta x = x_1 - x_0$. É preciso, também, um modo de denotar o valor da função $f(x)$ em vários valores de x . Na prática, o padrão é utilizar a notação $f(x_i)$ para representar o valor de $f(x)$ quando $x = x_i$. Assim, para a função $f(x) = 5 + x^2$, temos $f(0) = 5 + 0^2 = 5$; e, de modo semelhante, $f(2) = 5 + 2^2 = 9$ etc.

Quando x muda de um valor inicial x_0 para um novo valor $(x_0 + \Delta x)$, o valor da função $y = f(x)$ muda de $f(x_0)$ para $f(x_0 + \Delta x)$. A variação em y por unidade de variação em x pode ser representada pelo *quociente de diferenças*.

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \quad (6.1)$$

Esse quociente, que mede a taxa média de variação de y , pode ser calculado se conhecermos o valor inicial de x , ou x_0 , e a grandeza da variação em x , ou Δx . Isto é, $\Delta y/\Delta x$ é uma função de x_0 e Δx .

EXEMPLO 1 Dado $y = f(x) = 3x^2 - 4$, podemos escrever

$$f(x_0) = 3(x_0)^2 - 4 \quad f(x_0 + \Delta x) = 3(x_0 + \Delta x)^2 - 4$$

Portanto, o quociente de diferenças é

$$\begin{aligned} \frac{\Delta y}{\Delta x} &= \frac{3(x_0 + \Delta x)^2 - 4 - (3x_0^2 - 4)}{\Delta x} = \frac{6x_0\Delta x + 3(\Delta x)^2}{\Delta x} \\ &= 6x_0 + 3\Delta x \end{aligned} \quad (6.2)$$

que pode ser avaliado se tivermos x_0 e Δx . Seja $x_0 = 3$ e $\Delta x = 4$; então, a taxa de variação média de y é $6(3) + 3(4) = 30$. Isso significa que, na média, quando x muda de 3 para 7, a variação em y é de 30 unidades por unidade de variação em x .

A derivada

Freqüentemente estamos interessados na taxa de variação de y quando Δx é muito pequena. Nesse caso, é possível obter uma aproximação de $\Delta y/\Delta x$ descartando todos os termos do quociente de diferenças que envolvam a expressão Δx . Em (6.2), por exemplo, se Δx for muito pequena, podemos simplesmente tomar o termo $6x_0$ à direita como uma aproximação de $\Delta y/\Delta x$. Quanto menor for o valor de Δx , é claro, mais perto a aproximação ficará do verdadeiro valor de $\Delta y/\Delta x$.

À medida que Δx tende a zero (o que significa que fica cada vez mais próximo de zero, mas nunca o alcança, realmente), $(6x_0 + 3\Delta x)$ tenderá ao valor $6x_0$ e, pelo mesmo critério, $\Delta y/\Delta x$ também tenderá a $6x_0$. Esse fato é expresso simbolicamente ou pela afirmação $\Delta y/\Delta x \rightarrow 6x_0$ quando $\Delta x \rightarrow 0$, ou pela equação

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} (6x_0 + 3\Delta x) = 6x_0 \quad (6.3)$$

onde o símbolo $\lim_{\Delta x \rightarrow 0}$ lê-se “o limite de ... quando Δx tende a 0.” Se, quando $\Delta x \rightarrow 0$, o limite do quociente de diferenças $\Delta y/\Delta x$ realmente existir, esse limite é denominado a derivada da função $y = f(x)$.

Há vários pontos que devem ser notados em relação à derivada, se ela existir. Em primeiro lugar, uma derivada é uma *função*; de fato, conforme essa utilização, a palavra *derivada* realmente significa uma função derivada. A função original $y = f(x)$ é uma *função primitiva*, e a derivada é uma outra função derivada dela. Conquanto o quociente de diferenças seja uma função de x_0 e Δx , você deve observar – em (6.3), por exemplo – que a derivada é uma função somente de x_0 . Isso porque Δx já é obrigada a tender a zero e, portanto, não deve ser considerada como uma outra variável da função. Vamos acrescentar, também, que, até aqui, utilizamos o símbolo com índice x_0 apenas para frisar o fato de que uma variação em x deve se iniciar a partir de algum valor específico de x . Agora que isso está entendido, podemos eliminar o índice e declarar, simplesmente, que a derivada, como a função primitiva, é, em si, uma função da variável independente x . Isto é, para cada valor de x , há um único valor correspondente para a função derivada.

Em segundo lugar, já que a derivada é meramente um limite do quociente de diferenças que mede a taxa de variação de y , a derivada deve, necessariamente, também ser uma medida de uma taxa de variação. Em vista do fato de que a variação em x contemplada no conceito de derivada é infinitesimal (isto é, $\Delta x \rightarrow 0$), a taxa medida pela derivada tem as características de uma taxa de variação *instantânea*.

Em terceiro lugar, há uma questão de notação. Em geral, funções derivadas são comumente denotadas de duas maneiras. Dada uma função primitiva $y = f(x)$, um modo de denotar sua derivada (se ela existir) é utilizar o símbolo $f'(x)$, ou simplesmente f' ; essa notação é atribuída ao matemático Lagrange. A outra notação comum é dy/dx , inventada pelo matemático Leibniz. [Na verdade, há uma terceira notação, Dy , ou $Df(x)$, mas ela não será utilizada na discussão a seguir.] A notação $f'(x)$, que é parecida com a notação da função primitiva $f(x)$, tem a vantagem de transmitir a idéia de que a derivada é, em si, uma função de x . A razão para expressá-la como $f'(x)$ – em vez de, digamos, $\phi(x)$ – é enfatizar que a função f' é derivada da função primitiva f . A notação alternativa, dy/dx , ao contrário, serve para enfatizar que o valor de uma derivada mede uma taxa de variação. A letra d é a contraparte da letra grega Δ , e dy/dx é diferente de $\Delta y/\Delta x$ principalmente porque a primeira é o limite da última quando Δx tende a zero. Na discussão subsequente, utilizaremos ambas as notações, dependendo do que parecer mais conveniente em um contexto específico.

Utilizando essas duas notações, podemos definir a derivada de uma dada função $y = f(x)$ como se segue:

$$\frac{dy}{dx} = f'(x) \equiv \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$$

EXEMPLO 2

Mais uma vez, com referência à função $y = 3x^2 - 4$, mostraremos que seu quociente de diferenças é (6.2), e que o limite desse quociente é (6.3). Com base nesse último fato, podemos escrever (substituindo x_0 por x):

$$\frac{dy}{dx} = 6x \quad \text{ou} \quad f'(x) = 6x$$

Note que valores diferentes de x darão à derivada valores diferentes correspondentes. Por exemplo, quando $x = 3$, achamos, substituindo $x = 3$ na expressão $f'(x)$ que, $f'(3) = 6(3) = 18$; de modo semelhante, quando $x = 4$, temos $f'(4) = 6(4) = 24$. Assim, enquanto $f'(x)$ denota uma *função derivada*, cada uma das expressões $f'(3)$ e $f'(4)$ representa um valor específico da derivada.

EXERCÍCIO 6.2

- Dada a função $y = 4x^2 + 9$:
 - Calcule o quociente de diferenças como uma função de x e Δx . (Use x em lugar de x_0 .)
 - Calcule a derivada dy/dx .
 - Calcule $f'(3)$ e $f'(4)$.
- Dada a função $y = 5x^2 - 4x$:
 - Calcule o quociente de diferenças como uma função de x e Δx .
 - Calcule a derivada dy/dx .
 - Calcule $f'(2)$ e $f'(3)$.
- Dada a função $y = 5x - 2$:
 - Calcule o quociente de diferenças $\Delta y/\Delta x$. Que tipo de função ele é?
 - Visto que a expressão Δx não aparece na função $\Delta y/\Delta x$ na parte (a), faz alguma diferença para o valor de $\Delta y/\Delta x$ se Δx for grande ou pequeno? Por consequência, qual é o limite do quociente de diferenças quando Δx tende a zero?

6.3 A derivada e a inclinação de uma curva

A economia elementar nos diz que, dada uma função de custo total $C = f(Q)$, onde C denota o custo total e Q a produção, o custo marginal (MC) é definido como a variação no custo total resultante do aumento de uma unidade de produção; isto é, $MC = \Delta C/\Delta Q$. Entende-se que ΔQ é uma variação extremamente pequena. No caso de um produto em unidades discretas (inteiros somente), uma variação de uma unidade é a menor variação possível; mas, no caso de um produto cuja quantidade é uma variável contínua, ΔQ pode se referir a uma variação infinitesimal. Neste último caso, sabe-se muito bem que o custo marginal pode ser medido pela inclinação da curva do custo total. Mas a inclinação da curva do custo total nada mais é que o limite da razão $\Delta C/\Delta Q$, quando ΔQ tende a zero. Assim, o conceito da inclinação de uma curva é meramente a contrapartida geométrica do conceito de derivada. Ambos têm a ver com a noção de “marginal” utilizada tão extensamente na economia.

Na Figura 6.1, desenhamos uma curva de custo total C , que é a representação gráfica da função (primitiva) $C = f(Q)$. Suponha que consideramos Q_0 o nível inicial de produção, a partir do qual é medido um aumento de produção; então, o ponto relevante na curva de custo é o ponto A . Se a produção for aumentada para $Q_0 + \Delta Q = Q_2$, o custo total aumentará de C_0 para $C_0 + \Delta C = C_2$; assim $\Delta C/\Delta Q = (C_2 - C_0)/(Q_2 - Q_0)$. Em termos geométricos, essa é a razão entre dois segmentos de reta, EB/AE , ou a *inclinação* da reta AB . Essa razão particular mede uma taxa média de variação – o custo marginal *médio* para a ΔQ específica ilustrada – e representa um quociente de diferenças. Como tal, é uma função do valor inicial Q_0 e da quantidade de variação ΔQ .

O que acontece quando variamos a grandeza de ΔQ ? Se contemplarmos um incremento de produção menor (digamos, de Q_0 para Q_1 somente), então o custo marginal médio será medido pela inclinação da reta AD . Além disso, à medida que reduzimos o incremento de produção cada vez mais, resultarão retas cada vez menos inclinadas até que, no limite (quando $\Delta Q \rightarrow 0$), obtemos a reta KG (que é a *reta tangente* à curva de custo no ponto A) como a reta relevante. A inclinação de KG ($= HG/KH$) mede a inclinação da curva do custo total no ponto A e representa o limite de $\Delta C/\Delta Q$, quando $\Delta Q \rightarrow 0$, quando a produção inicial estiver em $Q = Q_0$. Portanto, em termos da derivada, a inclinação da curva $C = f(Q)$ no ponto A corresponde ao valor específico da derivada $f'(Q_0)$.

O limite à esquerda de q é simbolizado por $\lim_{v \rightarrow N^-} q$ (o sinal de menos significa que parte de valores menores que N), e o limite à direita é representado por $\lim_{v \rightarrow N^+} q$. Quando – e somente quando – os dois limites tiverem um valor finito comum (digamos, L), consideramos que o limite de q existe e o representamos como $\lim_{v \rightarrow N} q = L$. Note que L tem de ser um número *finito*. Se tivermos a situação de $\lim_{v \rightarrow N} q = \infty$ (ou $-\infty$), consideraremos que q não possui *nenhum* limite porque $\lim_{v \rightarrow N} q = \infty$ significa que $q \rightarrow \infty$ quando $v \rightarrow N$, e, se q assumirá valores *cada vez maiores* quando v tender a N , é contraditório dizer que q tem um limite. Entretanto, adotando um modo conveniente de expressar o fato de que $q \rightarrow \infty$ quando $v \rightarrow N$, há quem realmente escreva $\lim_{v \rightarrow N} q = \infty$ e afirme que q tem um “limite infinito”.

Em certos casos, somente o limite de um lado precisa ser considerado. Ao tomar o limite de q quando $v \rightarrow +\infty$, por exemplo, somente o limite à esquerda de q é relevante, porque v só pode tender a $+\infty$ pela esquerda. De modo semelhante, no caso de $v \rightarrow -\infty$, somente o limite à direita é relevante. Nesses casos, a existência ou não do limite de q depende apenas de q tender a um valor finito quando $v \rightarrow +\infty$ ou quando $v \rightarrow -\infty$.

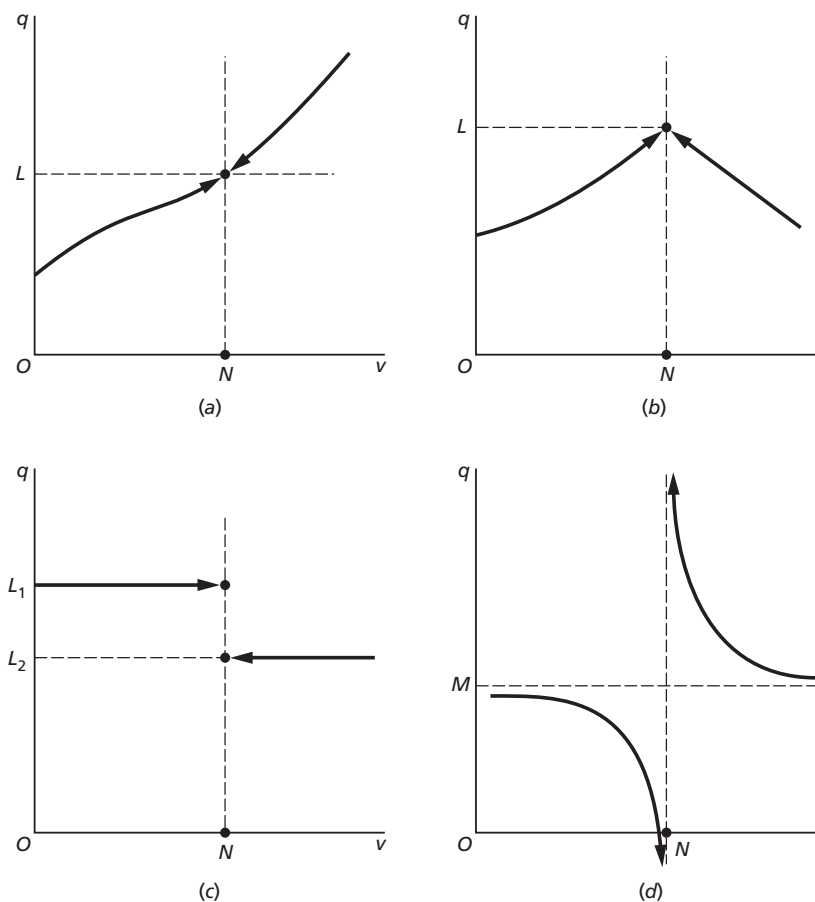
É importante perceber que o símbolo ∞ (infinito) não é um número e, portanto, não pode ser sujeito às operações algébricas usuais. Não podemos ter $3 + \infty$ nem $1/\infty$; nem podemos escrever $q = \infty$, que não é o mesmo que $q \rightarrow \infty$. Contudo, é aceitável expressar o *limite* de q como “=” (em vez de $\rightarrow$) ∞ , pois isso apenas indica que $q \rightarrow \infty$.

Ilustrações gráficas

Na Figura 6.2, vamos ilustrar diversas situações possíveis relativas ao limite de uma função $q = g(v)$.

A Figura 6.2a mostra uma curva suave. À medida que a variável v tende ao valor de N por *qualquer* lado no eixo horizontal, a variável q tende ao valor L . Nesse caso, o limite à direita é idêntico ao limite à esquerda; portanto, podemos escrever $\lim_{v \rightarrow N} q = L$.

FIGURA 6.2



A curva desenhada na Figura 6.2b não é suave; ela tem um ponto de inflexão localizado diretamente acima do ponto N . Não obstante, quando v tende a N de qualquer lado, q novamente tende a um valor idêntico L . O limite de q mais uma vez existe e é igual a L .

A Figura 6.2c mostra uma função conhecida como *escalonada*.[†] Nesse caso, quando v tende a N , o limite de q à esquerda é L_1 , mas o limite à direita é L_2 , um número diferente. Por conseguinte, q não tem um limite quando $v \rightarrow N$.

Por fim, na Figura 6.2d, quando v tende a N , o limite à esquerda de q é $-\infty$, enquanto o limite à direita é $+\infty$, porque as duas partes da curva (hiperbólica) cairão e subirão indefinidamente ao tenderem à reta vertical tracejada como uma assíntota. Novamente, $\lim_{v \rightarrow N} q$ não existe. Por outro lado, se estivermos considerando um tipo diferente de limite no diagrama d, a saber, $\lim_{v \rightarrow +\infty} q$, então somente o limite à esquerda tem relevância e verificamos que esse limite realmente existe:

$\lim_{v \rightarrow +\infty} q = M$. Analogamente, é possível verificar que $\lim_{v \rightarrow -\infty} q = M$ também.

Também é possível aplicar os conceitos de limites à direita e à esquerda à discussão do custo marginal na Figura 6.1. Naquele contexto, as variáveis q e v se referem, respectivamente, ao quociente $\Delta C/\Delta Q$ e à grandeza de ΔQ , e todas as variações são medidas a partir do ponto A na curva. Em outras palavras, q se refere à inclinação de retas tais como AB , AD e KG , enquanto v se refere ao comprimento de retas tais como Q_0Q_2 (= reta AE) e Q_0Q_1 (= reta AF). Já vimos que, quando v tende a zero a partir de um valor positivo, q tenderá a um valor igual à inclinação da reta KG . De modo semelhante, podemos estabelecer que, se ΔQ tender a zero a partir de um valor negativo (isto é, à medida que a *redução* da produção ficar cada vez menor), o quociente $\Delta C/\Delta Q$ medido pela inclinação de retas tais como RA (não desenhada) também tenderá a um valor igual à inclinação da reta KG . Na verdade, aqui, a situação é muito mais semelhante àquela ilustrada na Figura 6.2a. Assim, a inclinação de KG na Figura 6.1 (a contraparte de L na Figura 6.2) é, de fato, o limite do quociente q quando v tende a zero, e, como tal, nos dá o custo marginal no nível de produção $Q = Q_0$.

Avaliação de um limite

Vamos, agora, ilustrar a avaliação algébrica de um limite de uma dada função $q = g(v)$.

EXEMPLO 1 Dado $q = 2 + v^2$, encontre $\lim_{v \rightarrow 0} q$. Para tomar o limite à esquerda, substituímos v pela série de valores negativos $-1, -\frac{1}{10}, -\frac{1}{100}, \dots$ (nessa ordem) e achamos que $(2 + v^2)$ decrescerá regularmente e tenderá a 2 (porque v^2 gradativamente tenderá a 0). Em seguida, para o limite à direita, substituímos v pela série de valores positivos $1, \frac{1}{10}, \frac{1}{100}, \dots$ (nessa ordem) e achamos o mesmo limite encontrado anteriormente. Visto que os dois limites são idênticos, consideramos que o limite de q existe e escrevemos $\lim_{v \rightarrow 0} q = 2$.

É tentador considerar a resposta obtida no Exemplo 1 como o resultado de ter-se estabelecido $v = 0$ na equação $q = 2 + v^2$, mas, em geral, é preciso resistir a essa tentação. Ao avaliar $\lim_{v \rightarrow N} q$, deixamos apenas v tender a N , mas, como regra, não deixamos que $v = N$. De fato, podemos falar, com bastante legitimidade, do limite de q quando $v \rightarrow N$, mesmo quando N não estiver no domí-

[†] Esse nome pode ser facilmente explicado pelo formato da curva. Mas funções escalonadas também podem ser expressas em termos algébricos. A função ilustrada na Figura 6.2c pode ser expressa pela equação

$$q = \begin{cases} L_1 & (\text{para } 0 \leq v < N) \\ L_2 & (\text{para } N \leq v) \end{cases}$$

Note que, tal como descrito, em cada subconjunto de seu domínio, a função aparece como uma função constante distinta, que constitui um "degrau" no gráfico.

Em economia, funções escalonadas podem ser utilizadas, por exemplo, para mostrar os vários preços cobrados por diferentes quantidades compradas (a curva mostrada na Figura 6.2c ilustra o *desconto por quantidade*) ou as várias taxas de juros aplicadas a diferentes segmentos de renda.

nio da função $q = g(v)$. Neste último caso, se tentarmos estabelecer $v = N$, q claramente será indefinida.

EXEMPLO 2

Dado $q = (1 - v^2)/(1 - v)$, ache $\lim_{v \rightarrow 1} q$. Aqui, $N = 1$ não está no domínio da função e não podemos estabelecer $v = 1$ porque isso envolveria divisão por zero. Além do mais, mesmo o procedimento de avaliação de limite que obriga $v \rightarrow 1$, como utilizado no Exemplo 1, causará dificuldade, pois o denominador $(1 - v)$ tenderá a zero quando $v \rightarrow 1$ e continuaremos sem ter nenhum meio de efetuar a divisão no limite.

Um modo de escapar dessa dificuldade é tentar transformar a razão dada em uma forma na qual v não apareça no denominador. Posto que $v \rightarrow 1$ implica que $v \neq 1$, de modo que $(1 - v)$ é não-zero, é legítimo dividir a expressão $(1 - v^2)$ por $(1 - v)$, e escrever[†]

$$q = \frac{1 - v^2}{1 - v} = 1 + v \quad (v \neq 1)$$

Nessa nova expressão para q , não há mais um denominador contendo v . Visto que $(1 + v) \rightarrow 2$ quando $v \rightarrow 1$ de *qualquer* lado, podemos então concluir que $\lim_{v \rightarrow 1} q = 2$.

EXEMPLO 3

Dado $q = (2v + 5)/(v + 1)$, ache $\lim_{v \rightarrow +\infty} q$. A variável v novamente aparece no numerador e *também*

no denominador. Se deixarmos $v \rightarrow +\infty$ em ambos, o resultado será uma razão entre dois números infinitamente grandes, o que não tem um significado muito claro. Para escapar dessa dificuldade, desta vez tentamos transformar a razão dada em uma forma na qual a variável v não apareça no numerador.[‡] Mais uma vez, isso pode ser conseguido dividindo a razão dada. Visto que $(2v + 5)$ não é divisível exatamente por $(v + 1)$, contudo, o resultado conterá um termo de resto, como se segue:

$$q = \frac{2v + 5}{v + 1} = 2 + \frac{3}{v + 1}$$

Porém, de qualquer forma, essa nova expressão para q já não tem mais um numerador contendo v . Observando que o resto $3/(v + 1) \rightarrow 0$ quando $v \rightarrow +\infty$, podemos então concluir que $\lim_{v \rightarrow +\infty} q = 2$.

Também existem diversos teoremas úteis para a avaliação de limites, que serão discutidos na Seção 6.6.

Visão formal do conceito de limite

A discussão anterior deve ter transmitido algumas idéias gerais sobre o conceito de limite. Agora, daremos a ele uma definição mais precisa. Visto que essa definição fará uso do conceito de *vizi-*

[†] A divisão pode ser efetuada, como no caso de números, da seguinte maneira:

$$\begin{array}{r} 1 - v \quad \overline{) 1 - v^2} \\ \underline{1 - v} \\ 0 \end{array}$$

Alternativamente, podemos recorrer à fatoração, como segue:

$$\frac{1 - v^2}{1 - v} = \frac{(1 + v)(1 - v)}{1 - v} = 1 + v \quad (v \neq 1)$$

[‡] Note que, diferente do caso $v \rightarrow 0$, onde queremos retirar v do *denominador* para evitar divisão por zero, o caso $v \rightarrow \infty$ é melhor evitado retirando v do *numerador*. Quando $v \rightarrow \infty$, uma expressão que contenha v no numerador se torna infinita, mas uma expressão com v no denominador tende a zero e sai silenciosamente de cena, o que é mais conveniente para nós.

nhança de um ponto sobre uma reta (em particular, um número específico tal como um ponto sobre uma reta de números reais), em primeiro lugar explicaremos o último termo.

Para um dado número L , pode sempre ser encontrado um número $(L - a_1) < L$ e um outro número $(L + a_2) > L$, onde a_1 e a_2 são números positivos arbitrários. O conjunto de todos os números que se encontram entre $(L - a_1)$ e $(L + a_2)$ é denominado o *intervalo* entre esses dois números. Se os números $(L - a_1)$ e $(L + a_2)$ estiverem incluídos no conjunto, ele é um *intervalo fechado*; se estiverem excluídos, o conjunto é um *intervalo aberto*. Um intervalo fechado entre $(L - a_1)$ e $(L + a_2)$ é denotado pela expressão entre colchetes

$$[L - a_1, L + a_2] \equiv \{q \mid L - a_1 \leq q \leq L + a_2\}$$

e o intervalo *aberto* correspondente é denotado entre parênteses:

$$(L - a_1, L + a_2) \equiv \{q \mid L - a_1 < q < L + a_2\} \quad (6.4)$$

Assim, $[]$ está relacionado com o sinal de desigualdade fraca $\leq$, enquanto $()$ está relacionado com o sinal de desigualdade restrita $<$. Mas, em ambos os tipos de intervalos, o número menor $(L - a_1)$ sempre aparece em primeiro lugar. Mais adiante, teremos a oportunidade de ver também intervalos *meio-abertos* e *meio-fechados* como $(3, 5]$ e $[6, \infty)$, que têm os seguintes significados:

$$(3, 5] \equiv \{x \mid 3 < x \leq 5\} \quad [6, \infty) \equiv \{x \mid 6 \leq x < \infty\}$$

Agora podemos definir uma *vizinhança* de L como um intervalo aberto como definido em (6.4), que é um intervalo que “contém” o número L .[†] Dependendo das grandezas dos números arbitrários a_1 e a_2 , é possível construir várias vizinhanças para o dado número L . Usando o conceito de vizinhança, o limite de uma função pode, então, ser definido da seguinte maneira:

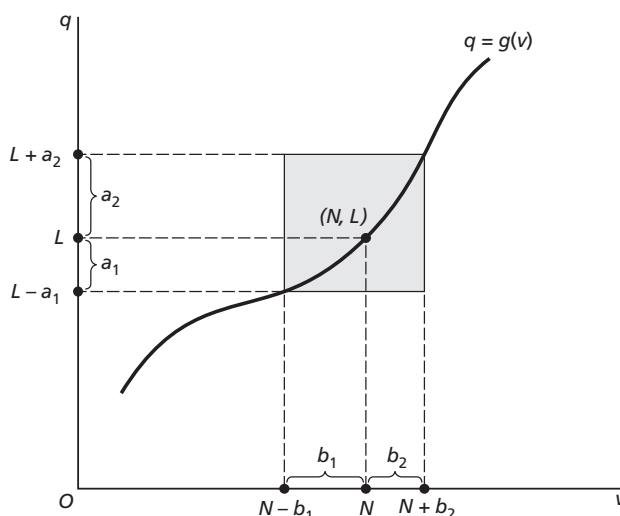
Quando v tende a um número N , o limite de $q = g(v)$ é o número L se, para toda vizinhança de L que puder ser escolhida, *não importa quão pequena*, for possível achar uma vizinhança correspondente de N (excluindo o ponto $v = N$) no domínio da função tal que, para todo valor de v naquela vizinhança de N , sua imagem esteja na vizinhança escolhida de L .

Esse enunciado pode ser esclarecido com a ajuda da Figura 6.3, que é parecida com a Figura 6.2a. Com o que já aprendemos com a Figura 6.2a, sabemos que $\lim_{v \rightarrow N} q = L$ na Figura 6.3. Agora, vamos mostrar que L realmente cumpre a nova definição de limite. Como primeira etapa, selecionamos arbitrariamente uma vizinhança pequena de L , digamos, $(L - a_1, L + a_2)$. (Essa diferença poderia ser até menor, mas a estamos mantendo relativamente grande para facilitar a explanação). Agora, construa uma vizinhança de N , digamos, $(N - b_1, N + b_2)$ tal que as duas vizinhanças (quando estendidas para o quadrante I) definirão, em conjunto, um retângulo (sombreado no diagrama) do qual duas de suas arestas estão sobre a curva dada. Pode-se verificar, então, que, para todo valor de v nessa vizinhança de N (sem contar $v = N$), o valor correspondente de $q = g(v)$ está na vizinhança escolhida de L . De fato, não importa quão *pequena* seja a vizinhança de L que escolhermos, pode-se achar uma vizinhança de N (suficientemente pequena) com a propriedade que acabamos de citar. Assim, L cumpre a definição de um limite, como queríamos demonstrar.

Podemos, também, aplicar a definição dada à função escalonada da Figura 6.2c para mostrar que nem L_1 nem L_2 se qualificam como $\lim_{v \rightarrow N} q$. Se escolhermos uma vizinhança muito pequena de L_1 – digamos, à distância de apenas um fio de cabelo de cada lado de L_1 –, então, não importa qual vizinhança escolhermos para N , o retângulo associado com as duas vizinhanças não tem nenhuma possibilidade de conter o degrau inferior da função. Por consequência, para qualquer valor de $v > N$, o valor correspondente de q (localizado no degrau inferior) não estará na vizinhan-

[†] A identificação de um intervalo aberto como a vizinhança de um ponto é válida somente quando estivermos considerando um ponto sobre uma reta (um espaço unidimensional). No caso de um ponto sobre um plano (espaço bidimensional), sua vizinhança pode ser imaginada como uma área, por exemplo, uma área circular que inclui o ponto.

FIGURA 6.3



ça de L_1 e, assim, L_1 não passa no teste para um limite. Por um raciocínio similar, L_2 também deve ser descartada como candidata para $\lim_{v \rightarrow N} q$. De fato, nesse caso, não existe nenhum limite para q quando $v \rightarrow N$.

O cumprimento da definição também pode ser verificado algebricamente, em vez de por meios gráficos. Por exemplo, considere, mais uma vez, a função

$$q = \frac{1-v^2}{1-v} = 1+v \quad (v \neq 1) \quad (6.5)$$

Vimos, no Exemplo 2, que $\lim_{v \rightarrow 1} q = 2$; assim, aqui temos $N=1$ e $L=2$. Para verificar que $L=2$ é, de fato, o limite de q , temos de demonstrar que, para toda vizinhança escolhida de L , $(2-a_1, 2+a_2)$, existe uma vizinhança escolhida de N , $(1-b_1, 1+b_2)$, tal que, sempre que v estiver nessa vizinhança de N , q deve estar na vizinhança escolhida de L . Isso significa, em essência, que, para dados valores de a_1 e a_2 , não importa quão pequenos devem ser encontrados dois números b_1 e b_2 tais que, sempre que a desigualdade

$$1-b_1 < v < 1+b_2 \quad (v \neq 1) \quad (6.6)$$

for satisfeita, uma outra desigualdade, da forma

$$2-a_1 < q < 2+a_2 \quad (6.7)$$

também deve ser satisfeita. Para achar tal par de números b_1 e b_2 , vamos primeiramente reescrever (6.7) substituindo por (6.5):

$$2-a_1 < 1+v < 2+a_2 \quad (6.7')$$

Essa expressão pode, por sua vez, ser transformada (subtraindo 1 de cada lado) na desigualdade

$$1-a_1 < v < 1+a_2 \quad (6.7'')$$

Comparar (6.7'') – uma variante de (6.7) – com (6.6) sugere que, se escolhermos os dois números b_1 e b_2 de modo que $b_1 = a_1$ e $b_2 = a_2$, as duas desigualdades (6.6) e (6.7) sempre serão satisfeitas simultaneamente. Assim, a vizinhança de N , $(1-b_1, 1+b_2)$, pode, de fato, ser achada para o caso de $L=2$, como exige a definição de um limite, e isso estabelece $L=2$ como o limite.

Vamos agora inverter a definição de um limite e utilizá-la para mostrar que um outro valor, (digamos, 3) não pode se qualificar como $\lim_{v \rightarrow 1} q$ para a função em (6.5). Se 3 fosse aquele limite, teria de ser verdade que, para toda vizinhança escolhida de 3, $(3 - a_1, 3 + a_2)$, existe uma vizinhança de 1, $(1 - b_1, 1 + b_2)$, tal que, sempre que v estiver nesta última vizinhança, q deve estar na vizinhança anterior. Isto é, sempre que a desigualdade

$$1 - b_1 < v < 1 + b_2$$

for satisfeita, uma outra desigualdade da forma

$$3 - a_1 < 1 + v < 3 + a_2$$

ou

$$2 - a_1 < v < 2 + a_2$$

também deve ser satisfeita. O *único* modo de conseguir esse resultado é escolher $b_1 = a_1 - 1$ e $b_2 = a_2 + 1$, o que implicaria que a vizinhança de 1 deve ser o intervalo aberto $(2 - a_1, 2 + a_2)$. Segundo a definição de limite, contudo, a_1 e a_2 podem ser arbitrariamente pequenos, digamos, $a_1 = a_2 = 0,1$. Nesse caso, o intervalo mencionado por último será $(1,9, 2,1)$, que está inteiramente à direita do ponto $v = 1$ no eixo horizontal e, portanto, nem ao menos se qualifica como uma vizinhança de 1. Assim, a definição de um limite não pode ser satisfeita pelo número 3. Um procedimento semelhante pode ser empregado para demonstrar que *qualquer* número que não seja 2 contradiz a definição de um limite no presente caso.

Em geral, se um número satisfizer a definição de um limite de q quando $v \rightarrow N$, então nenhum outro número poderá satisfazê-la. Se existir um limite, ele é único.

EXERCÍCIO 6.4

1. Dada a função $q = (v^2 + v - 56)/(v - 7)$, ($v \neq 7$), calcule o limite à esquerda e o limite à direita de q quando v tende a 7. Com essas respostas, podemos concluir que q tem um limite quando v tende a 7?
2. Dado $q = [(v + 2)^3 - 8]/v$, ($v \neq 0$), calcule:

(a) $\lim_{v \rightarrow 0} q$	(b) $\lim_{v \rightarrow 2} q$	(c) $\lim_{v \rightarrow a} q$
--------------------------------	--------------------------------	--------------------------------
3. Dado $q = 5 - 1/v$, ($v \neq 0$), calcule:

(a) $\lim_{v \rightarrow +\infty} q$	(b) $\lim_{v \rightarrow -\infty} q$
--------------------------------------	--------------------------------------
4. Use a Figura 6.3 para mostrar que *não podemos* considerar o número $(L + a_2)$ como o limite de q quando v tende a N .

6.5 Digressão sobre desigualdades e valores absolutos

Encontramos sinais de desigualdade muitas vezes antes. Conforme exposto da Seção 6.4, também aplicamos operações matemáticas a desigualdades. Ao transformar (6.7') em (6.7''), por exemplo, subtraímos 1 de cada lado da desigualdade. Quais são as regras de operações que se podem aplicar de modo geral a desigualdades (em vez de a equações)?

Regras de desigualdades

Para começar, vamos enunciar uma importante propriedade das desigualdades: elas são *transitivas*. Isso significa que, se $a > b$ e se $b > c$, então $a > c$. Visto que igualdades (equações) também são transitivas, a propriedade de transitividade deve se aplicar às desigualdades “fracas” ($\geq$ ou $\leq$), bem como às “estritas” ($>$ ou $<$). Assim, temos

$$a > b, b > c \Rightarrow a > c$$

$$a \geq b, b \geq c \Rightarrow a \geq c$$

Essa propriedade é o que torna possível escrever uma *desigualdade contínua*, tal como $3 < a < b < 8$ ou $7 \leq x \leq 24$. (Como regra, ao escrever uma desigualdade contínua, os sinais de desigualdade são arranjados na mesma direção, usualmente com o menor número à esquerda.)

As regras mais importantes de desigualdades são as que governam a adição (subtração) de um número a (de) uma desigualdade, a multiplicação ou divisão de uma desigualdade por um número e a elevação de uma desigualdade ao quadrado. Especificamente, essas regras são as seguintes.

Regra I (adição e subtração) $a > b \Rightarrow a \pm k > b \pm k$

Uma desigualdade continuará válida se uma quantidade igual for adicionada ou subtraída de cada lado. Essa regra pode ser generalizada assim: se $a > b > c$, então $a \pm k > b \pm k > c \pm k$.

Regra II (multiplicação e divisão)

$$a > b \Rightarrow \begin{cases} ka > kb & (k > 0) \\ ka < kb & (k < 0) \end{cases}$$

A multiplicação de ambos os lados por um número *positivo* preserva a desigualdade, mas um multiplicador *negativo* causará a inversão do *sentido* (ou *direção*) da desigualdade.

EXEMPLO 1 Uma vez que $6 > 5$, a multiplicação por 3 resultará em $3(6) > 3(5)$, ou $18 > 15$; mas a multiplicação por -3 resultará em $(-3)6 < (-3)5$, ou $-18 < -15$.

A divisão de uma desigualdade por um número n é equivalente à multiplicação pelo número $1/n$; portanto, a regra de divisão é incorporada pela regra de multiplicação.

Regra III (elevação ao quadrado) $a > b, (b \geq 0) \Rightarrow a^2 > b^2$

Se ambos os lados forem não-negativos, a desigualdade continuará a valer quando os dois forem elevados ao quadrado.

EXEMPLO 2 Visto que $4 > 3$ e visto que ambos os lados são positivos, temos $4^2 > 3^2$, ou $16 > 9$. De modo semelhante, visto que $2 > 0$, segue-se que $2^2 > 0^2$, ou $4 > 0$.

As regras de I a III foram enunciadas em termos de desigualdades estritas, mas sua validade não é afetada se os sinais $>$ forem substituídos por sinais $\geq$.

Valores absolutos e desigualdades

Quando o domínio de uma variável x é um intervalo aberto (a, b) , o domínio pode ser denotado pelo conjunto $\{x \mid a < x < b\}$ ou, mais simplesmente, pela desigualdade $a < x < b$. De modo semelhante, se for um intervalo fechado $[a, b]$, ele pode ser expresso pela desigualdade fraca $a \leq x \leq b$. No caso especial de um intervalo da forma $(-a, a)$ – digamos, $(-10, 10)$ – ele pode ser representado ou pela desigualdade $-10 < x < 10$ ou, alternativamente, pela desigualdade

$$|x| < 10$$

onde o símbolo $|x|$ denota o *valor absoluto* (ou *valor numérico*) de x .

Para qualquer número real n , o valor absoluto de n é definido como se segue:[†]

$$|n| \equiv \begin{cases} n & (\text{se } n > 0) \\ -n & (\text{se } n < 0) \\ 0 & (\text{se } n = 0) \end{cases} \quad (6.8)$$

Note que, se $n = 15$, então $|15| = 15$; mas se $n = -15$, temos

$$|-15| = -(-15) = 15$$

também. Portanto, o valor absoluto de qualquer número real é, com efeito, simplesmente seu valor numérico após a remoção do sinal. Por essa razão, sempre temos $|n| = |-n|$. O valor absoluto de n também é denominado o *módulo* de n .

Dada a expressão $|x| = 10$, podemos concluir, de (6.8), que x pode ser ou 10 ou -10 . Pelo mesmo critério, a expressão $|x| < 10$ significa que (1) se $x > 0$, então $x \equiv |x| < 10$, de modo que x deve ser menor que 10; mas também que (2) se $x < 0$, então, de acordo com (6.8), temos $-x \equiv |x| < 10$, ou $x > -10$, de modo que x deve ser maior que -10 . Por conseguinte, combinando as duas partes desse resultado, vemos que x deve estar dentro do intervalo aberto $(-10, 10)$. Em geral, podemos escrever

$$|x| < n \Leftrightarrow -n < x < n \quad (n > 0) \quad (6.9)$$

que também pode ser estendido a desigualdades fracas como segue:

$$|x| \leq n \Leftrightarrow -n \leq x \leq n \quad (n \geq 0) \quad (6.10)$$

Porque são, em si, números, os valores absolutos de dois números m e n podem ser somados, subtraídos, multiplicados e divididos. As seguintes propriedades caracterizam valores absolutos:

$$|m| + |n| \geq |m + n|$$

$$|m| \cdot |n| = |m \cdot n|$$

$$\frac{|m|}{|n|} = \left| \frac{m}{n} \right|$$

O interessante é que a primeira delas envolve uma desigualdade, e não uma equação. É fácil perceber a razão para isso: enquanto a expressão à esquerda $|m| + |n|$ é, definitivamente, uma *soma* de dois valores numéricos (ambos tomados como positivos), a expressão $|m + n|$ é o valor numérico *ou* de uma soma (se m e n forem, digamos, ambos positivos) *ou* de uma diferença (se m e n tiverem sinais opostos). Assim, o lado esquerdo pode ser maior que o lado direito.

EXEMPLO 3

Se $m = 5$ e $n = 3$, então $|m| + |n| = |m + n| = 8$. Mas, se $m = 5$ e $n = -3$, então $|m| + |n| = 5 + 3 = 8$, enquanto

$$|m + n| = |5 - 3| = 2$$

é um número menor.

Nas outras duas propriedades, por outro lado, não faz diferença se m e n tiverem sinais idênticos ou opostos, já que, ao tomar o valor absoluto do produto ou do quociente do lado direito, de qualquer modo o sinal do último termo será removido.

[†] Mais uma vez, alertamos que, embora a notação de valor absoluto seja semelhante à de um determinante de primeira ordem, esses dois conceitos são inteiramente diferentes. A definição de um determinante de primeira ordem é $|a_{ij}| \equiv a_{ij}$, independentemente do sinal de a_{ij} . Na definição do valor absoluto $|n|$, por outro lado, o sinal de n fará diferença. O contexto da discussão normalmente deve deixar claro se o que está sendo considerado é um valor absoluto ou um determinante de primeira ordem.

EXEMPLO 4 Se $m = 7$ e $n = 8$, então $|m| \cdot |n| = |m \cdot n| = 7(8) = 56$. Mas, mesmo se $m = -7$ e $n = 8$ (sinais opostos), ainda obtemos o mesmo resultado de

$$|m| \cdot |n| = |-7| \cdot |8| = 7(8) = 56$$

e $|m \cdot n| = |-7(8)| = 7(8) = 56$

Solução de uma desigualdade

Assim como uma equação, uma desigualdade que contém uma variável (digamos, x) pode ter uma solução; a solução, se existir, é um conjunto de valores de x que fazem da desigualdade uma afirmação verdadeira. Essa solução será, em si, usualmente na forma de uma desigualdade.

EXEMPLO 5 Calcule a solução da desigualdade

$$3x - 3 > x + 1$$

Como na resolução de uma equação, em primeiro lugar os termos variáveis devem ser reunidos de um lado da desigualdade. Adicionando $(3 - x)$ a ambos os lados, obtemos

$$3x - 3 + 3 - x > x + 1 + 3 - x$$

ou

$$2x > 4$$

Multiplicando ambos os lados por $\frac{1}{2}$ (o que não inverte o sentido da desigualdade porque $\frac{1}{2} > 0$), teremos a solução

$$x > 2$$

que é, em si, uma desigualdade. Essa solução não é um número único, mas um conjunto de números. Por conseguinte, também podemos expressar a solução como o conjunto $\{x \mid x > 2\}$ ou como o intervalo aberto $(2, \infty)$.

EXEMPLO 6 Resolva a desigualdade $|1 - x| \leq 3$. Em primeiro lugar, vamos nos livrar da notação de valor absoluto utilizando (6.10). A desigualdade dada é equivalente a afirmar que

$$-3 \leq 1 - x \leq 3$$

ou, após subtrair 1 de cada lado,

$$-4 \leq -x \leq 2$$

Multiplicando cada lado por (-1) , obtemos, então,

$$4 \geq x \geq -2$$

onde o sentido da desigualdade foi devidamente invertido. Escrevendo o número menor em primeiro lugar, podemos expressar a solução na forma da desigualdade

$$-2 \leq x \leq 4$$

ou na forma do conjunto $\{x \mid -2 \leq x \leq 4\}$ ou do intervalo fechado $[-2, 4]$.

Às vezes, um problema pode exigir que diversas desigualdades sejam satisfeitas com diversas variáveis simultaneamente; então, temos de resolver um sistema de desigualdades simultâneas. Esse problema surge, por exemplo, em programação não-linear, que será discutida no Capítulo 13.

EXERCÍCIO 6.5

- Resolva as seguintes desigualdades:

(a) $3x - 1 < 7x + 2$	(c) $5x + 1 < x + 3$
(b) $2x + 5 < x - 4$	(d) $2x - 1 < 6x + 5$
- Se $8x - 3 < 0$ e $8x > 0$, expresse essas desigualdades como uma desigualdade contínua e encontre a solução.
- Resolva as seguintes expressões:

(a) $ x + 1 < 6$	(b) $ 4 - 3x < 2$	(c) $ 2x + 3 \leq 5$
-------------------	--------------------	-----------------------

6.6 Teoremas de limite

Nosso interesse em taxas de variação nos levou a considerar o conceito de derivada que, por ter as características do limite de um quociente de diferenças, nos sugeriu, por sua vez, o estudo das questões da existência e da avaliação de um limite. O processo básico de avaliação de limites, ilustrado na Seção 6.4, envolve deixar a variável v tender a um número específico (digamos, N) e observar o valor ao qual q tende. Entretanto, quando estivermos realmente avaliando o limite de uma função, podemos lançar mão de certos teoremas estabelecidos sobre limites, que podem simplificar a tarefa na prática, especialmente no caso de funções complicadas.

Teoremas que envolvem uma única função

Quando estiver envolvida uma única função, $q = g(v)$, podemos aplicar os seguintes teoremas.

Teorema I Se $q = av + b$, então $\lim_{v \rightarrow N} q = aN + b$ (a e b são constantes).

EXEMPLO 1 Dado $q = 5v + 7$, temos $\lim_{v \rightarrow 2} q = 5(2) + 7 = 17$. Da mesma forma, $\lim_{v \rightarrow 0} q = 5(0) + 7 = 7$.

Teorema II Se $q = g(v) = b$, então $\lim_{v \rightarrow N} q = b$.

Esse teorema, que diz que o limite de uma função constante é a constante naquela função, é um mero caso especial do Teorema I, quando $a = 0$. (Você já viu um exemplo desse caso no Exercício 6.2-3.)

Teorema III Se $q = v$, então $\lim_{v \rightarrow N} q = N$.
 Se $q = v^k$, então $\lim_{v \rightarrow N} q = N^k$.

EXEMPLO 2 Dado $q = v^3$, temos $\lim_{v \rightarrow 2} q = (2)^3 = 8$.

Você talvez tenha notado que o que fazemos nos teoremas I até III para achar o limite de q quando $v \rightarrow N$ é, na realidade, deixar que $v = N$. Mas esses casos são especiais e não invalidam a regra geral que afirma que “ $v \rightarrow N$ ” não significa “ $v = N$.”

Teoremas que envolvem duas funções

Se tivermos duas funções da mesma variável independente v , $q_1 = g(v)$ e $q_2 = h(v)$ e se ambas as funções possuem limites como segue:

$$\lim_{v \rightarrow N} q_1 = L_1 \quad \lim_{v \rightarrow N} q_2 = L_2$$

onde L_1 e L_2 são dois números *finitos*, os seguintes teoremas podem ser aplicados.

Teorema IV (teorema do limite da soma/diferença)

$$\lim_{v \rightarrow N} (q_1 \pm q_2) = L_1 \pm L_2$$

O limite de uma soma (diferença) de duas funções é a soma (diferença) de seus respectivos limites.

Em particular, notamos que

$$\lim_{v \rightarrow N} 2q_1 = \lim_{v \rightarrow N} (q_1 + q_1) = L_1 + L_1 = 2L_1$$

que está de acordo com o Teorema I.

Teorema V (teorema do limite do produto)

$$\lim_{v \rightarrow N} (q_1 q_2) = L_1 L_2$$

O limite de um produto de duas funções é o produto de seus limites.

Aplicado ao quadrado de uma função, isso resulta em

$$\lim_{v \rightarrow N} (q_1 q_1) = L_1 L_1 = L_1^2$$

que está de acordo com o Teorema III.

Teorema VI (teorema do limite do quociente)

$$\lim_{v \rightarrow N} \frac{q_1}{q_2} = \frac{L_1}{L_2} \quad (L_2 \neq 0)$$

O limite de um quociente de duas funções é o quociente de seus limites. Naturalmente, o limite L_2 deve ser não-zero; caso contrário, o quociente é indefinido.

EXEMPLO 3 Calcule $\lim_{v \rightarrow 0} (1 + v)/(2 + v)$. Visto que temos aqui $\lim_{v \rightarrow 0} (1 + v) = 1$ e $\lim_{v \rightarrow 0} (2 + v) = 2$, o limite desejado é $\frac{1}{2}$.

Lembre-se de que L_1 e L_2 representam números finitos; caso contrário, esses teoremas não se aplicariam. Além do mais, no caso do Teorema VI, L_2 também tem de ser não-zero. Se essas restrições não forem satisfeitas, temos de recorrer ao método de avaliação de limites ilustrado nos exemplos 2 e 3 na Seção 6.4, relacionados, respectivamente, aos casos em que L_2 é zero e L_2 é infinito.

Limite de uma função polinomial

Tendo à nossa disposição os teoremas de limite dados, podemos avaliar com facilidade o limite de qualquer função polinomial

$$q = g(v) = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n \quad (6.11)$$

quando v tende ao número N . Posto que os limites dos termos separados são, respectivamente,

$$\lim_{v \rightarrow N} a_0 = a_0 \quad \lim_{v \rightarrow N} a_1 v = a_1 N \quad \lim_{v \rightarrow N} a_2 v^2 = a_2 N^2 \quad (\text{etc.})$$

o limite da função polinomial é (pelo teorema do limite da soma)

$$\lim_{v \rightarrow N} q = a_0 + a_1 N + a_2 N^2 + \cdots + a_n N^n \quad (6.12)$$

Notamos que, na verdade, esse limite também é igual a $g(N)$, isto é, é igual ao valor da função em (6.11) quando $v = N$. Esse resultado específico será importante quando discutirmos o conceito de *continuidade* da função polinomial.

EXERCÍCIO 6.6

1. Calcule os limites da função $q = 7 - 9v + v^2$:
 - (a) Quando $v \rightarrow 0$
 - (b) Quando $v \rightarrow 3$
 - (c) Quando $v \rightarrow -1$
2. Calcule os limites de $q = (v + 2)(v - 3)$:
 - (a) Quando $v \rightarrow -1$
 - (b) Quando $v \rightarrow 0$
 - (c) Quando $v \rightarrow 5$
3. Calcule os limites de $q = (3v + 5)/(v + 2)$:
 - (a) Quando $v \rightarrow 0$
 - (b) Quando $v \rightarrow 5$
 - (c) Quando $v \rightarrow -1$

6.7 Continuidade e diferenciabilidade de uma função

A discussão precedente sobre o conceito de limite e sua avaliação agora pode ser usada para definir a continuidade e a diferenciabilidade de uma função. Essas noções afetam diretamente a derivada da função, que é o que nos interessa.

Continuidade de uma função

Quando uma função $q = g(v)$ possui um limite quando v tende ao ponto N no domínio, e quando esse limite também é igual a $g(N)$ – isto é, ao valor da função em $v = N$ – diz-se que a função é *contínua* em N . O termo *continuidade*, do modo como é definido aqui, envolve nada menos que três requisitos: (1) o ponto N deve estar no domínio da função, isto é, $g(N)$ é definida; (2) a função deve ter um limite quando $v \rightarrow N$; isto é, $\lim_{v \rightarrow N} g(v)$ existe; e (3) o limite deve ter valor igual a $g(N)$, isto é, $\lim_{v \rightarrow N} g(v) = g(N)$.

É importante notar que, conquanto o ponto (N, L) tenha sido excluído de consideração quando discutimos o limite da curva na Figura 6.3, no presente contexto não o estamos mais excluindo. Ao contrário, como o terceiro requisito determina especificamente, o ponto (N, L) deve estar no gráfico da função antes que a função possa ser considerada contínua no ponto N .

Vamos verificar se as funções mostradas na Figura 6.2 são contínuas. No diagrama *a*, todos os três requisitos são cumpridos no ponto N . O ponto N está no domínio; q tem o limite L quando $v \rightarrow N$; e o limite L também é, por acaso, o valor da função em N . Assim, a função representada por aquela curva é contínua em N . O mesmo é verdade para a função ilustrada na Figura 6.2*b*, visto que L é o limite da função quando v tende ao valor N no domínio e, já que L é também o valor da função em N . Este último exemplo gráfico deve ser suficiente para estabelecer que a continuidade de uma função no ponto N não implica necessariamente que o gráfico da função será “suave” em $v = N$, pois o ponto (N, L) na Figura 6.2*b* é, na verdade, um ponto de forte “inflexão” e, ainda assim, a função é contínua naquele valor de v .

Quando uma função $q = g(v)$ é contínua para todos os valores de v no intervalo (a, b) , diz-se que ela é contínua naquele intervalo. Se a função for contínua em todos os pontos de um subconjunto S do domínio (onde o subconjunto S pode ser a união de diversos intervalos disjuntos), diz-se que ela é contínua em S . E, por fim, se a função for contínua em todos os pontos de seu domínio, dizemos que ela é contínua em seu domínio. Mesmo neste último caso, todavia, o gráfico da função pode, não obstante, revelar uma descontinuidade (uma lacuna) em algum valor de v , digamos, em $v = 5$, se o valor de v não estiver em seu domínio.

Referindo-nos mais uma vez à Figura 6.2, vemos que, no diagrama *c*, a função é *descontínua* em N porque não existe um limite naquele ponto, o que viola o segundo requisito de continuidade.

de. Não obstante, a função satisfaz os requisitos de continuidade no intervalo $(0, N)$ do domínio, bem como no intervalo $[N, \infty)$. O diagrama *d* também é obviamente descontínuo em $v = N$. Desta vez, a descontinuidade resulta do fato de que N é excluído do domínio, violando o primeiro requisito de continuidade.

Com base nos gráficos na Figura 6.2, parece que pontos de inflexão são consistentes com continuidade, como no diagrama *b*, mas que lacunas são tabu, como nos diagramas *c* e *d*. E é esse realmente o caso. Portanto, em termos simples, uma função que é contínua em um intervalo específico é uma função cujo gráfico pode ser desenhado para o intervalo citado sem levantar o lápis ou a caneta do papel – o que é possível mesmo quando há pontos de inflexão, mas impossível quando ocorrem lacunas.

Funções polinomiais e racionais

Vamos considerar, agora, a continuidade de certas funções que encontramos freqüentemente. Para qualquer função polinomial, tal como $q = g(v)$ em (6.11), (6.12) implica que $\lim_{v \rightarrow N} q$ existe e é igual ao valor da função em N . Posto que N é um ponto (qualquer ponto) no domínio de uma função, podemos concluir que qualquer função polinomial é contínua em seu domínio. Essa é uma informação muito útil, porque encontramos funções polinomiais com muita freqüência.

E as funções racionais? Em relação à continuidade, existe um teorema interessante (o teorema da continuidade) que estabelece que a soma, a diferença, o produto e o quociente de qualquer número finito de funções que são contínuas no domínio são, respectivamente, também contínuos no domínio. O resultado é que qualquer função racional (um quociente de duas funções polinomiais) também deve ser contínua em seu domínio.

EXEMPLO 1 A função racional

$$q = g(v) = \frac{4v^2}{v^2 + 1}$$

é definida para todos os números reais finitos; assim, seu domínio consiste do intervalo $(-\infty, \infty)$. Para qualquer número N no domínio, o limite de q é (pelo teorema do limite do quociente)

$$\lim_{v \rightarrow N} q = \frac{\lim_{v \rightarrow N} (4v^2)}{\lim_{v \rightarrow N} (v^2 + 1)} = \frac{4N^2}{N^2 + 1}$$

que é igual a $g(N)$. Assim, todos os três requisitos de continuidade são satisfeitos em N . Além do mais, notamos que N pode representar qualquer ponto no domínio dessa função; por consequência, essa função é contínua em seu domínio.

EXEMPLO 2 A função racional

$$q = \frac{v^3 + v^2 - 4v - 4}{v^2 - 4}$$

não é definida em $v = 2$ e em $v = -2$. Visto que esses dois valores de v não estão no domínio, a função é descontínua em $v = -2$ e $v = 2$, a despeito do fato de existir um limite de q quando $v \rightarrow -2$ ou 2 . Graficamente, essa função apresentará uma lacuna em cada um desses dois valores de v . Mas, para outros valores de v (os que estão no domínio), essa função é contínua.

Diferenciabilidade de uma função

A discussão anterior nos deu as ferramentas para nos certificarmos se qualquer função tem um limite quando sua variável independente tende a algum valor específico. Assim, podemos tentar tomar o limite de qualquer função $y = f(x)$ quando x tende a algum valor escolhido, digamos, x_0 . Contudo, também podemos aplicar o conceito de “limite” em um nível diferente e tomar o limite do quociente de diferenças daquela função, $\Delta y / \Delta x$, quando Δx tende a zero. Os resultados de

tomar limites nesses dois níveis diferentes dizem respeito a duas propriedades diferentes, embora relacionadas, da função f .

Tomando o limite da função $y = f(x)$ em si, de acordo com o que discutimos na subseção precedente, podemos examinar se a função f é *contínua* em $x = x_0$. As condições para continuidade são: (1) $x = x_0$ deve estar no domínio da função f ; (2) y deve ter um limite quando $x \rightarrow x_0$; e (3) o limite em questão deve ser igual a $f(x_0)$. Quando esses requisitos são satisfeitos, podemos escrever

$$\lim_{x \rightarrow x_0} f(x) = f(x_0) \quad [\text{condição de continuidade}] \quad (6.13)$$

Ao contrário, quando o conceito de “limite” é aplicado ao quociente de diferenças $\Delta y / \Delta x$ quando $\Delta x \rightarrow 0$, estamos lidando com a questão de uma função ser ou não ser *diferenciável* em $x = x_0$, isto é, se a derivada dy/dx existe em $x = x_0$, ou se $f'(x_0)$ existe. O termo *diferenciável* é usado aqui porque o processo de obtenção da derivada dy/dx é conhecido como *diferenciação* (também denominado *derivação*). Posto que $f'(x_0)$ existe se e somente se o limite de $\Delta y / \Delta x$ existir em $x = x_0$ quando $\Delta x \rightarrow 0$, a expressão simbólica da diferenciabilidade de f é

$$\begin{aligned} f'(x_0) &= \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \\ &\equiv \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \quad [\text{condição de diferenciabilidade}] \end{aligned} \quad (6.14)$$

Essas duas propriedades, continuidade e diferenciabilidade, estão intimamente relacionadas – a continuidade de f é uma condição *necessária* para sua diferenciabilidade (embora, como veremos mais adiante, essa condição *não é suficiente*). Isso significa que, para ser diferenciável em $x = x_0$, a função deve, em primeiro lugar, passar no teste de ser contínua em $x = x_0$. Para provar isso, demonstraremos que, dada uma função $y = f(x)$, sua continuidade em $x = x_0$ resulta de sua diferenciabilidade em $x = x_0$; isto é, a condição (6.13) resulta da condição (6.14). Contudo, antes de fazer isso, vamos simplificar um pouco a notação (1) substituindo x_0 pelo símbolo N e (2) substituindo $(x_0 + \Delta x)$ pelo símbolo x . Este último é justificável porque a variação posterior do valor de x pode ser qualquer número (dependendo da grandeza da variação) e, por conseguinte, é uma variável que pode ser denotada por x . A equivalência dos dois sistemas de notação é mostrada na Figura 6.4, onde as notações antigas aparecem (dentro de colchetes) ao lado das novas. Note que, com a mudança de notação, Δx agora se torna $(x - N)$, de modo que a expressão “ $\Delta x \rightarrow 0$ ” se torna “ $x \rightarrow N$ ”, que é análoga à expressão $v \rightarrow N$ usada anteriormente em relação à função $q = g(v)$. De acordo com isso, (6.13) e (6.14) agora podem ser reescritas, respectivamente, como

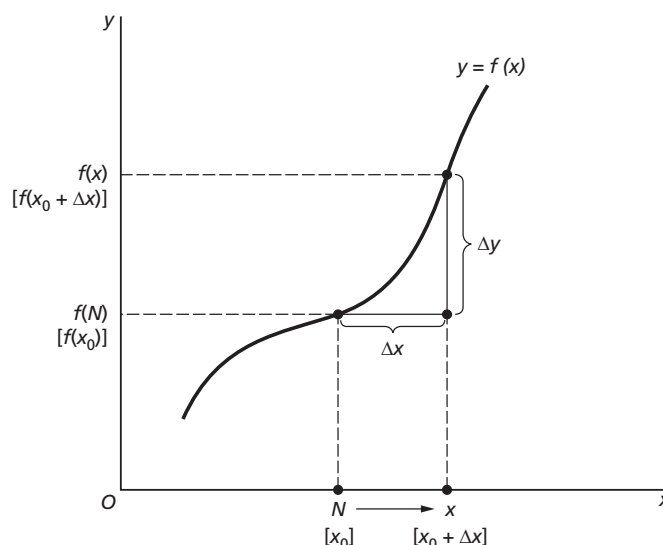
$$\lim_{x \rightarrow N} f(x) = f(N) \quad (6.13')$$

$$f'(N) = \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} \quad (6.14')$$

O que queremos mostrar é, portanto, que a condição de continuidade (6.13') resulta da condição de diferenciabilidade (6.14'). Primeiro, visto que a notação $x \rightarrow N$ implica que $x \neq N$, de modo que $x - N$ é um número não-zero, é permissível escrever a seguinte identidade:

$$f(x) - f(N) \equiv \frac{f(x) - f(N)}{x - N} (x - N) \quad (6.15)$$

FIGURA 6.4



Tomando o limite de cada lado de (6.15) quando $x \rightarrow N$, temos os seguintes resultados:

$$\begin{aligned}
 \text{Lado esquerdo} &= \lim_{x \rightarrow N} f(x) - \lim_{x \rightarrow N} f(N) && [\text{teorema do limite da diferença}] \\
 &= \lim_{x \rightarrow N} f(x) - f(N) && [f(N) \text{ é uma constante}] \\
 \text{Lado direito} &= \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} \lim_{x \rightarrow N} (x - N) && [\text{teorema do limite do produto}] \\
 &= f'(N) (\lim_{x \rightarrow N} x - \lim_{x \rightarrow N} N) && [\text{por (6.14')} \text{ e teorema do limite da diferença}] \\
 &= f'(N) (N - N) = 0
 \end{aligned}$$

Note que não poderíamos ter escrito esses resultados se a condição (6.14') não tivesse sido levada em conta, pois, se $f'(N)$ não existisse, a expressão do lado direito (e, por conseguinte, também a do lado esquerdo) em (6.15) não possuiria um limite. Contudo, se $f'(N)$ existir, os dois lados terão limites, como mostrado nas equações anteriores. Além disso, quando se igualam os resultados do lado direito e do lado esquerdo, obtemos $\lim_{x \rightarrow N} f(x) - f(N) = 0$, que é idêntico a (6.13').

Assim, provamos que continuidade, como mostra em (6.13'), resulta da diferenciabilidade, como mostra (6.14'). Em geral, se uma função é diferenciável em todos os pontos em seu domínio, podemos concluir que ela deve ser contínua em seu domínio.

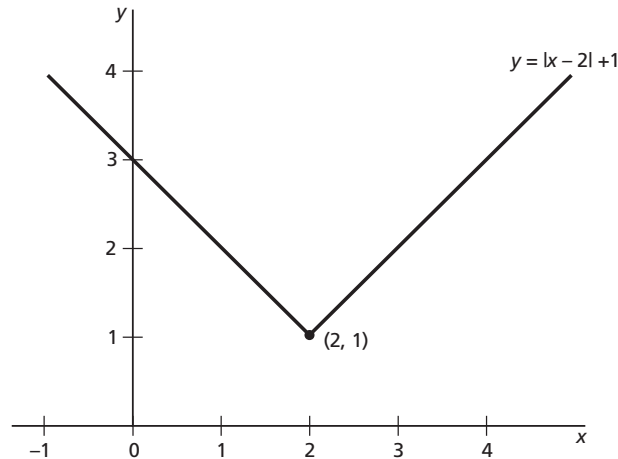
Embora a diferenciabilidade implique continuidade, o contrário não é verdade. Ou seja, a continuidade é uma condição *necessária*, mas *não suficiente* para a diferenciabilidade. Para demonstrar isso, basta produzir um contra-exemplo. Vamos considerar a função

$$y = f(x) = |x - 2| + 1 \quad (6.16)$$

apresentada graficamente na Figura 6.5. Como pode-se mostrar de pronto, essa função não é diferenciável, embora seja contínua, quando $x = 2$. Determinar que a função é contínua em $x = 2$ é fácil. Primeiro, $x = 2$ está no domínio da função. Segundo, o limite de y existe quando x tende a 2; especificamente, $\lim_{x \rightarrow 2^-} y = \lim_{x \rightarrow 2^+} y = 1$. Terceiro, também achamos que $f(2) = 1$. Assim, todos os três requisitos de continuidade são cumpridos. Para mostrar que a função f não é diferenciável em $x = 2$, devemos mostrar que o limite do quociente de diferenças

$$\lim_{x \rightarrow 2} \frac{f(x) - f(2)}{x - 2} = \lim_{x \rightarrow 2} \frac{|x - 2| + 1 - 1}{x - 2} = \lim_{x \rightarrow 2} \frac{|x - 2|}{x - 2}$$

FIGURA 6.5



não existe. Isso envolve a demonstração de uma disparidade entre os limites do lado esquerdo e do lado direito. Uma vez que, ao considerar o limite do lado direito, x deve ser maior que 2 segundo a definição de valor absoluto em (6.8), temos $|x - 2| = x - 2$. Assim, o limite do lado direito é

$$\lim_{x \rightarrow 2^+} \frac{|x - 2|}{x - 2} = \lim_{x \rightarrow 2^+} \frac{x - 2}{x - 2} = \lim_{x \rightarrow 2^+} 1 = 1$$

Por outro lado, ao considerar o limite do lado esquerdo, x deve ser menor que 2; assim, segundo (6.8), $|x - 2| = -(x - 2)$. Por consequência, o limite do lado esquerdo é

$$\lim_{x \rightarrow 2^-} \frac{|x - 2|}{x - 2} = \lim_{x \rightarrow 2^-} \frac{-(x - 2)}{x - 2} = \lim_{x \rightarrow 2^-} (-1) = -1$$

que é diferente do limite do lado direito. Isso mostra que a continuidade não garante a diferenciabilidade. Em suma, todas as funções diferenciáveis são contínuas, mas nem todas as funções contínuas são diferenciáveis.

Na Figura 6.5, a não-diferenciabilidade da função em $x = 2$ é evidente pelo fato de o ponto $(2, 1)$ não ter nenhuma reta tangente definida e, por conseguinte, não ser possível atribuir nenhuma inclinação definida ao ponto. Especificamente, à esquerda daquele ponto a curva tem uma inclinação de -1 , mas à direita, tem uma inclinação de $+1$, e as inclinações dos dois lados não apresentam nenhuma propensão a tender a uma grandeza comum em $x = 2$. É claro que o ponto $(2, 1)$ é um ponto especial; é o único ponto de inflexão na curva. Em outros pontos da curva, a derivada é definida e a função é diferenciável. Mais especificamente, a função em (6.16) pode ser dividida em duas funções lineares como segue:

$$\text{Parte esquerda:} \quad y = -(x - 2) + 1 = 3 - x \quad (x \leq 2)$$

$$\text{Parte direita:} \quad y = (x - 2) + 1 = x - 1 \quad (x > 2)$$

A parte esquerda é diferenciável no intervalo $(-\infty, 2)$, e a parte direita é diferenciável no intervalo $(2, \infty)$ no domínio.

Em geral, a condição de diferenciabilidade é mais restritiva que a condição de continuidade, porque requer algo além da continuidade. A continuidade em um ponto somente exclui a presença de uma lacuna, enquanto a diferenciabilidade também exclui a “inflexão”. Portanto, diferenciabilidade exige “suavidade” da função (curva) bem como sua continuidade. Grande parte das funções *específicas* empregadas em economia têm a propriedade de ser diferenciáveis em toda a parte. Além do mais, quando são usadas funções *gerais*, freqüentemente admite-se que elas são diferenciáveis em toda a parte, o que faremos na discussão subsequente.

EXERCÍCIO 6.7

1. Uma função $y = f(x)$ é descontínua em $x = x_0$ quando *qualquer* dos três requisitos de continuidade for violado em $x = x_0$. Construa três gráficos para ilustrar a violação de cada um desses requisitos.
2. Tomando o conjunto de todos os números reais finitos como o domínio da função $q = g(v) = v^2 - 5v - 2$:
 - (a) Calcule o limite de q quando v tende a N (um número real finito).
 - (b) Verifique se esse limite é igual a $g(N)$.
 - (c) Verifique se a função é contínua em N e contínua em seu domínio.
3. Dada a função $q = g(v) = \frac{v+2}{v^2+2}$:
 - (a) Use os teoremas de limite para achar $\lim_{v \rightarrow N} q$, quando N for um número real finito.
 - (b) Verifique se esse limite é igual a $g(N)$.
 - (c) Verifique a continuidade da função $g(v)$ em N em seu domínio $(-\infty, \infty)$.
4. Dado $y = f(x) = \frac{x^2 - 9x + 20}{x - 4}$:
 - (a) É possível aplicar o teorema do limite do quociente para calcular o limite dessa função quando $x \rightarrow 4$?
 - (b) Essa função é contínua em $x = 4$? Por quê?
 - (c) Ache uma função que, para $x \neq 4$, é equivalente à função dada, e obtenha, a partir da função equivalente, o limite de y quando $x \rightarrow 4$.
5. Na função racional no Exemplo 2, o numerador é divisível exatamente pelo denominador e o quociente é $v + 1$. Dada essa razão, podemos substituir a função diretamente por $q = v + 1$? Justifique sua resposta.
6. Com base nos gráficos das seis funções na Figura 2.8, você concluiria que cada uma delas é diferenciável em todos os pontos em seu domínio? Explique.

CAPÍTULO 7

Regras de diferenciação e sua utilização em estática comparativa

O problema central da análise estática comparativa – achar uma taxa de variação – pode ser identificado com o problema de achar a derivada de uma função $y = f(x)$, quando se considera apenas uma variação infinitesimal em x . Ainda que a derivada dy/dx seja definida como o limite do quociente de diferenças $q = g(v)$ quando $v \rightarrow 0$, não é, de modo algum, necessário executar o processo de tomar o limite toda vez que procuramos a derivada de uma função, pois existem várias regras de diferenciação (derivação) que nos permitirão obter diretamente as derivadas desejadas. Portanto, em vez de partir imediatamente para modelos estáticos comparativos, vamos começar aprendendo algumas regras de diferenciação.

7.1 Regras de diferenciação para uma função de uma variável

Em primeiro lugar, vamos discutir três regras que se aplicam, respectivamente, aos seguintes tipos de funções de uma variável independente: $y = k$ (função constante) e $y = x^n$ e $y = cx^n$ (funções exponenciais). Todas essas funções apresentam gráficos suaves e contínuos e, portanto, são diferenciáveis em todos os pontos.

Regra da função constante

A derivada de uma função constante $y = k$, ou $f(x) = k$, é identicamente zero, isto é, é zero para todos os valores de x . Simbolicamente, esta regra pode ser enunciada como:

Dado $y = f(x) = k$, a derivada é:

$$\frac{dy}{dx} = \frac{dk}{dx} = 0 \quad \text{ou} \quad f'(x) = 0$$

Outra alternativa é enunciar a regra como: dado $y = f(x) = k$, a derivada é:

$$\frac{d}{dx} y = \frac{d}{dx} f(x) = \frac{d}{dx} k = 0$$

onde o símbolo de derivada foi dividido em duas partes, d/dx por um lado e y [ou $f(x)$ ou k] de outro. A primeira parte, d/dx , é um *símbolo operador*, que indica que devemos efetuar uma deter-

minada operação matemática. Exatamente como o símbolo operador $\sqrt{\quad}$ nos indica que devemos calcular uma raiz quadrada, o símbolo d/dx representa uma instrução para tomar a derivada ou diferenciar (alguma função) com relação à variável x . A função sobre a qual executaremos a operação (que será diferenciada) é indicada na segunda parte; aqui, ela é $y = f(x) = k$.

A demonstração da regra é a seguinte. Dada $f(x) = k$, temos $f(N) = k$ para qualquer valor de N . Assim, o valor de $f'(N)$ – o valor da derivada em $x = N$ – como definido em (6.13) é:

$$f'(N) = \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} = \lim_{x \rightarrow N} \frac{k - k}{x - N} = \lim_{x \rightarrow N} 0 = 0$$

Além disso, visto que N representa absolutamente qualquer valor de x , o resultado $f'(N) = 0$ pode ser imediatamente generalizado para $f'(x) = 0$, o que prova o resultado.

É importante distinguir claramente entre a afirmação $f'(x) = 0$ e a afirmação parecida, mas de significado diferente $f'(x_0) = 0$. Com $f'(x) = 0$, queremos dizer que a função derivada f' tem valor zero para *todos* os valores de x ; quando escrevemos $f'(x_0) = 0$, por outro lado, estamos meramente associando o valor zero da derivada e um valor particular de x , a saber, $x = x_0$.

Como já discutimos antes, a contraparte geométrica da derivada de uma função é a inclinação da curva. O gráfico de uma função constante, digamos, uma função de custo fixo $C_F = f(Q) = \$1.200$, é uma reta horizontal com inclinação zero em todos os seus pontos. Por correspondência, a derivada também deve ser zero para todos os valores de Q :

$$\frac{d}{dQ} C_F = \frac{d}{dQ} 1200 = 0$$

Regra da função de potência

A derivada de uma função de potência $y = f(x) = x^n$ é nx^{n-1} . Simbolicamente, isso é expresso como:

$$\frac{d}{dx} x^{n-1} = nx^{n-1} \quad \text{ou} \quad f'(x) = nx^{n-1} \quad (7.1)$$

EXEMPLO 1

A derivada de $y = x^3$ é $\frac{dy}{dx} = \frac{d}{dx} x^3 = 3x^2$.

EXEMPLO 2

A derivada de $y = x^9$ é $\frac{d}{dx} x^9 = 9x^8$.

Essa regra é válida para qualquer potência de x de valor real; isto é, o expoente pode ser qualquer número real. Mas, agora, provaremos a regra somente para o caso em que n é um inteiro positivo. No caso mais simples, quando $n = 1$, a função é $f(x) = x$ e, de acordo com a regra, a derivada é

$$f'(x) = \frac{d}{dx} x = 1(x^0) = 1$$

A demonstração desse resultado é obtida com facilidade a partir da definição de $f'(N)$ em (6.14'). Dada $f(x) = x$, o valor da derivada em qualquer valor de x , digamos, $x = N$, é:

$$f'(N) = \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} = \lim_{x \rightarrow N} \frac{x - N}{x - N} = \lim_{x \rightarrow N} 1 = 1$$

Posto que N representa qualquer valor de x , é permitido escrever $f'(x) = 1$. Isso prova a regra para o caso de $n = 1$. Como contraparte gráfica desse resultado, vemos que a função $y = f(x) = x$ é representada por uma reta a 45° , cuja inclinação é $+1$ em toda a sua extensão.

No caso de inteiros maiores, $n = 2, 3, \dots$, em primeiro lugar, vamos observar as seguintes identidades:

$$\begin{aligned} \frac{x^2 - N^2}{x - N} &= x + N && [2 \text{ termos à direita}] \\ \frac{x^3 - N^3}{x - N} &= x^2 + Nx + N^2 && [3 \text{ termos à direita}] \\ &\vdots \\ \frac{x^n - N^n}{x - N} &= x^{n-1} + Nx^{n-2} + N^2x^{n-3} + \dots + N^{n-1} && [n \text{ termos à direita}] \end{aligned} \quad (7.2)$$

Tomando (7.2) como base, podemos expressar a derivada de uma função de potência $f(x) = x^n$ em $x = N$ da seguinte maneira:

$$\begin{aligned} f'(N) &= \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} = \lim_{x \rightarrow N} \frac{x^n - N^n}{x - N} \\ &= \lim_{x \rightarrow N} (x^{n-1} + Nx^{n-2} + \dots + N^{n-1}) && [\text{por (7.2)}] \\ &= \lim_{x \rightarrow N} x^{n-1} + \lim_{x \rightarrow N} Nx^{n-2} + \dots + \lim_{x \rightarrow N} N^{n-1} && [\text{teorema do limite da soma}] \\ &= N^{n-1} + N^{n-1} + \dots + N^{n-1} && [\text{um total de } n \text{ termos}] \\ &= nN^{n-1} \end{aligned} \quad (7.3)$$

Novamente, N é qualquer valor de x ; assim, este último resultado pode ser generalizado para:

$$f'(x) = nx^{n-1}$$

que prova a regra para n , qualquer inteiro positivo.

Como mencionado anteriormente, essa regra se aplica mesmo quando o expoente n na expressão exponencial x^n não for um inteiro positivo. Os exemplos abaixo servem para ilustrar a aplicação da regra a estes últimos casos.

EXEMPLO 3 Calcule a derivada de $y = x^0$. Aplicando (7.1), encontramos:

$$\frac{d}{dx} x^0 = 0(x^{-1}) = 0$$

EXEMPLO 4 Calcule a derivada de $y = 1/x^3$. Esse caso envolve a recíproca de uma potência, mas, reescrevendo a função como $y = x^{-3}$, novamente podemos aplicar (7.1) para obter a derivada:

$$\frac{d}{dx} x^{-3} = -3x^{-4} \quad \left[= \frac{-3}{x^4} \right]$$

EXEMPLO 5 Calcule a derivada de $y = \sqrt{x}$. Neste caso, há uma raiz quadrada envolvida, mas, visto que $\sqrt{x} = x^{1/2}$, a derivada pode ser achada da seguinte maneira:

$$\frac{d}{dx} x^{1/2} = \frac{1}{2} x^{-1/2} \quad \left[= \frac{1}{2\sqrt{x}} = \frac{\sqrt{x}}{2x} \right]$$

Derivadas são, em si, funções da variável independente x . No Exemplo 1, por exemplo, a derivada é $dy/dx = 3x^2$, ou $f'(x) = 3x^2$, de modo que um valor diferente de x resultará em um valor diferente da derivada, tal como:

$$f'(1) = 3(1)^2 = 3 \quad f'(2) = 3(2)^2 = 12$$

Esses valores específicos da derivada podem ser expressos alternativamente como:

$$\left. \frac{dy}{dx} \right|_{x=1} = 3 \quad \left. \frac{dy}{dx} \right|_{x=2} = 12$$

mas as notações $f'(1)$ e $f'(2)$ são obviamente preferíveis, por causa de sua simplicidade.

É de suprema importância perceber que, para achar os valores das derivadas $f'(1)$, $f'(2)$ etc., em primeiro lugar temos de diferenciar a função $f(x)$ para obter a função derivada $f'(x)$, e, então, deixar que x assuma valores específicos em $f'(x)$. Substituir valores específicos de x na função primitiva $f(x)$ antes da diferenciação é terminantemente proibido. À guisa de ilustração, se admitirmos que $x = 1$ na função do Exemplo 1 antes da diferenciação, a função tomará o valor $y = x = 1$ – uma função constante –, que resultará em uma derivada zero em vez da resposta correta $f'(x) = 3x^2$.

Generalização da regra da função de potência

Quando uma constante multiplicativa c aparece na função potência, de modo que $f(x) = cx^n$, sua derivada é:

$$\frac{d}{dx} cx^n = cnx^{n-1} \quad \text{ou} \quad f'(x) = cnx^{n-1}$$

Esse resultado mostra que, ao diferenciar cx^n , podemos simplesmente manter intacta a constante multiplicativa c e então diferenciar o termo x^n de acordo com (7.1).

EXEMPLO 6 Dado $y = 2x$, temos $dy/dx = 2x^0 = 2$.

EXEMPLO 7 Dada $f(x) = 4x^3$, a derivada é $f'(x) = 12x^2$.

EXEMPLO 8 A derivada de $f(x) = 3x^{-2}$ é $f'(x) = -6x^{-3}$.

Para provar essa nova regra, considere o fato de que, para qualquer valor de x , digamos, $x = N$, o valor da derivada de $f(x) = cx^n$ é:

$$\begin{aligned} f'(N) &= \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} = \lim_{x \rightarrow N} \frac{cx^n - cN^n}{x - N} = \lim_{x \rightarrow N} c \left(\frac{x^n - N^n}{x - N} \right) \\ &= \lim_{x \rightarrow N} c \lim_{x \rightarrow N} \frac{x^n - N^n}{x - N} && \text{[teorema do limite do produto]} \\ &= c \lim_{x \rightarrow N} \frac{x^n - N^n}{x - N} && \text{[limite de uma constante]} \\ &= cnN^{n-1} && \text{[de (7.3)]} \end{aligned}$$

Uma vez que N é qualquer valor de x , este último resultado pode ser imediatamente generalizado para $f'(x) = cnx^{n-1}$, o que prova a regra.

EXERCÍCIO 7.1

- Calcule a derivada de cada uma das seguintes funções:
 (a) $y = x^{12}$ (c) $y = 7x^5$ (e) $w = -4u^{1/2}$
 (b) $y = 63$ (d) $w = 3u^{-1}$ (f) $w = 4u^{1/4}$
- Calcule:
 (a) $\frac{d}{dx}(-x^{-4})$ (c) $\frac{d}{dw}5w^4$ (e) $\frac{d}{du}au^b$
 (b) $\frac{d}{dx}9x^{1/3}$ (d) $\frac{d}{dx}cx^2$ (f) $\frac{d}{du}-au^{-b}$
- Calcule $f'(1)$ e $f'(2)$ das seguintes funções:
 (a) $y = f(x) = 18x$ (c) $f(x) = -5x^{-2}$ (e) $f(w) = 6w^{1/3}$
 (b) $y = f(x) = cx^3$ (d) $f(x) = \frac{3}{4}x^{4/3}$ (f) $f(w) = -3w^{-1/6}$
- Represente em gráfico a função $f(x)$ que dá origem à função derivada $f'(x) = 0$. Depois, represente em gráfico a função $g(x)$ caracterizada por $g'(x_0) = 0$.

7.2 Regras de diferenciação envolvendo duas ou mais funções da mesma variável

Cada uma das três regras apresentadas na Seção 7.1 refere-se a uma única função dada $f(x)$. Agora, suponha que temos duas funções *diferenciáveis* da mesma variável x , digamos, $f(x)$ e $g(x)$, e que queremos diferenciar a soma, a diferença, o produto ou o quociente formado com essas duas funções. Existem regras adequadas que se aplicam nessas circunstâncias? Ou, em termos mais concretos, dadas duas funções – digamos, $f(x) = 3x^2$ e $g(x) = 9x^{12}$ –, como obtemos a derivada de, por exemplo, $3x^2 + 9x^{12}$, ou a derivada de $(3x^2)(9x^{12})$?

Regra da soma/diferença

A derivada de uma soma (diferença) de duas funções é a soma (diferença) das derivadas das duas funções:

$$\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}f(x) \pm \frac{d}{dx}g(x) = f'(x) \pm g'(x)$$

Mais uma vez, a demonstração dessa regra envolve a aplicação da definição de derivada e dos vários teoremas de limite. Omitiremos a demonstração e, em seu lugar, apenas verificaremos a validade da regra e ilustraremos sua aplicação.

EXEMPLO 1

Dada a função $y = 14x^3$, dela podemos obter a derivada $dy/dx = 42x^2$. Mas $14x^3 = 5x^3 + 9x^3$, de modo que y pode ser considerada a soma de duas funções $f(x) = 5x^3$ e $g(x) = 9x^3$. Segundo a regra da soma, temos, então:

$$\frac{dy}{dx} = \frac{d}{dx}(5x^3 + 9x^3) = \frac{d}{dx}5x^3 + \frac{d}{dx}9x^3 = 15x^2 + 27x^2 = 42x^2$$

que é idêntico ao nosso resultado anterior.

Essa regra, que enunciamos em termos de duas funções, pode ser estendida com facilidade a mais funções. Assim, também é válido escrever:

$$\frac{d}{dx}[f(x) \pm g(x) \pm h(x)] = f'(x) \pm g'(x) \pm h'(x)$$

EXEMPLO 2 A função citada no Exemplo 1, $y = 14x^3$, pode ser escrita como $y = 2x^3 + 13x^3 - x^3$. A derivada da última, segundo a regra da soma/diferença, é:

$$\frac{dy}{dx} = \frac{d}{dx}(2x^3 + 13x^3 - x^3) = 6x^2 + 39x^2 - 3x^2 = 42x^2$$

que, novamente, está de acordo com nossa resposta anterior.

Essa regra tem grande importância na prática. Com ela à nossa disposição, agora é possível achar a derivada de qualquer função polinomial, visto que esta última nada mais é do que uma soma de funções exponenciais.

EXEMPLO 3

$$\frac{d}{dx}(ax^2 + bx + c) = 2ax + b$$

EXEMPLO 4

$$\frac{d}{dx}(7x^4 + 2x^3 - 3x + 37) = 28x^3 + 6x^2 - 3 + 0 = 28x^3 + 6x^2 - 3$$

Note que, nos Exemplos 3 e 4, as constantes c e 37 , na realidade, não têm nenhum efeito sobre a derivada, porque a derivada de um termo constante é zero. Ao contrário da constante *multiplicativa*, que é mantida durante a diferenciação, a constante *aditiva* é descartada. Esse fato proporciona a explanação matemática do famoso princípio econômico que diz que o custo fixo de uma empresa não afeta seu custo marginal. Dada uma função de custo total a curto prazo:

$$C = Q^3 - 4Q^2 + 10Q + 75$$

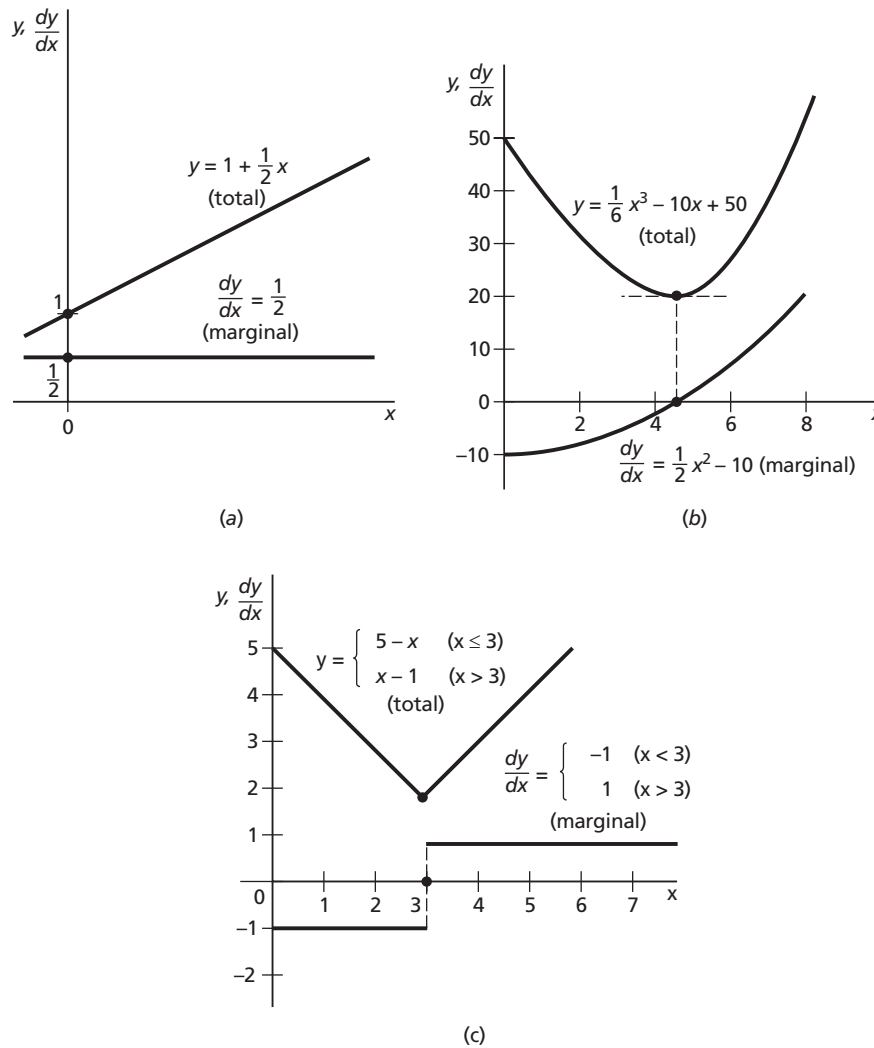
a função de custo marginal (para variação infinitesimal de produção) é o limite do quociente $\Delta C/\Delta Q$, ou a derivada da função C :

$$\frac{dC}{dQ} = 3Q^2 - 8Q + 10$$

enquanto o custo fixo é representado pela constante aditiva 75. Visto que a última é descartada durante o processo de derivação de dC/dQ , é óbvio que a grandeza do custo fixo não pode afetar o custo marginal.

Em geral, se uma função primitiva $y = f(x)$ representar uma função *total*, então a função derivada dy/dx é sua função *marginal*. É claro que ambas as funções podem ser representadas graficamente em relação à variável x ; e, por causa da correspondência entre a derivada de uma função e a inclinação de sua curva, para cada valor de x a função marginal deve mostrar a inclinação da função total naquele valor de x . Na Figura 7.1a, pode-se ver que uma função total linear (de inclinação constante) tem uma função marginal constante. Por outro lado, a função total não-linear (de inclinação variável) na Figura 7.1b dá origem a uma função marginal curva, que fica abaixo (acima) do eixo horizontal quando a curva da função total tem inclinação negativa (positiva). E, por fim, o leitor pode notar, na Figura 7.1c (cf. Figura 6.5), que a “não-suavidade” de uma função total resultará em uma lacuna (descontinuidade) na função marginal ou derivada. Isso contrasta acentuadamente com a função total suave em toda a sua extensão apresentada na Figura 7.1b, que dá origem a uma função marginal contínua. Por essa razão, a *suavidade* de uma função *primitiva* pode ser vinculada à *continuidade* de sua função *derivada*. Em particular, em vez de dizer que uma certa função é suave (e diferenciável) em toda a sua extensão, podemos caracterizá-la alter-

FIGURA 7.1



nativamente como uma função que tem uma função derivada contínua e nos referirmos a ela como uma função *continuamente diferenciável*.

As seguintes notações são muito utilizadas para denotar a continuidade e a diferenciabilidade contínua de uma função f :

$$f \in C^{(0)} \quad \text{ou} \quad f \in C: \quad f \text{ é contínua}$$

$$f \in C^{(1)} \quad \text{ou} \quad f \in C': \quad f \text{ é continuamente diferenciável}$$

onde $C^{(0)}$, ou simplesmente C , é o símbolo para o conjunto de todas as funções contínuas, e $C^{(1)}$, ou C' , é o símbolo do conjunto de todas as funções continuamente diferenciáveis.

Regra do produto

A derivada do produto de duas funções (diferenciáveis) é igual à primeira função vezes a derivada da segunda função mais a segunda função vezes a derivada da primeira função:

$$\begin{aligned} \frac{d}{dx} [f(x)g(x)] &= f(x) \frac{d}{dx} g(x) + g(x) \frac{d}{dx} f(x) \\ &= f(x)g'(x) + g(x)f'(x) \end{aligned} \quad (7.4)$$

É claro que também é possível rearranjar os termos e expressar a regra como:

$$\frac{d}{dx} [f(x)g(x)] = f'(x)g(x) + f(x)g'(x) \quad (7.4')$$

EXEMPLO 5 Calcule a derivada de $y = (2x + 3)(3x^2)$. Seja $f(x) = 2x + 3$ e $g(x) = 3x^2$. Então, segue-se que $f'(x) = 2$ e $g'(x) = 6x$, e, de acordo com (7.4), a derivada desejada é:

$$\frac{d}{dx} [(2x + 3)(3x^2)] = (2x + 3)(6x) + (3x^2)(2) = 18x^2 + 18x$$

Esse resultado pode ser verificado efetuando, em primeiro lugar, a multiplicação $f(x)g(x)$ e então tomando a derivada da polinomial do produto. A polinomial do produto é, nesse caso, $f(x)g(x) = (2x + 3)(3x^2) = 6x^3 + 9x^2$, e a diferenciação direta realmente dá a mesma derivada, $18x^2 + 18x$.

O ponto importante a lembrar é que a derivada de um produto de duas funções *não é* o simples produto das duas derivadas isoladas. É uma soma ponderada de $f'(x)$ e $g'(x)$, sendo os pesos, respectivamente, $g(x)$ e $f(x)$. Visto que isso é diferente do que a generalização intuitiva nos levaria a esperar, vamos produzir uma demonstração para (7.4). Conforme (6.13), o valor da derivada de $f(x)g(x)$ quando $x = N$ deve ser:

$$\left. \frac{d}{dx} [f(x)g(x)] \right|_{x=N} = \lim_{x \rightarrow N} \frac{f(x)g(x) - f(N)g(N)}{x - N} \quad (7.5)$$

Mas, somando e subtraindo $f(x)g(N)$ no numerador (o que não acarreta mudança na grandeza original), podemos transformar o quociente à direita de (7.5) como se segue:

$$\begin{aligned} & \frac{f(x)g(x) - f(x)g(N) + f(x)g(N) - f(N)g(N)}{x - N} \\ &= f(x) \frac{g(x) - g(N)}{x - N} + g(N) \frac{f(x) - f(N)}{x - N} \end{aligned}$$

Substituindo isso no quociente à direita de (7.5) e tomando seu limite, obtemos, então:

$$\begin{aligned} \left. \frac{d}{dx} [f(x)g(x)] \right|_{x=N} &= \lim_{x \rightarrow N} f(x) \lim_{x \rightarrow N} \frac{g(x) - g(N)}{x - N} \\ &+ \lim_{x \rightarrow N} g(N) \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} \end{aligned} \quad (7.5')$$

As quatro expressões de limite em (7.5') são facilmente avaliadas. A primeira é $f(N)$, e a terceira é $g(N)$ (limite de uma constante). As duas restantes são, de acordo com (6.13), respectivamente, $g'(N)$ e $f'(N)$. Assim, (7.5') se reduz a:

$$\left. \frac{d}{dx} [f(x)g(x)] \right|_{x=N} = f(N)g'(N) + g(N)f'(N) \quad (7.5'')$$

E, posto que N representa qualquer valor de x , (7.5'') continua válida se substituirmos cada símbolo N por x . Isso prova a regra.

Como extensão da regra para o caso de *três* funções, temos:

$$\begin{aligned} \frac{d}{dx} [f(x)g(x)h(x)] &= f'(x)g(x)h(x) + f(x)g'(x)h(x) \\ &+ f(x)g(x)h'(x) \quad [\text{conforme (7.4')}] \end{aligned} \quad (7.6)$$

Por extenso, a derivada do produto de três funções é igual ao produto da segunda e da terceira funções vezes a derivada da primeira, mais o produto da primeira e da terceira funções vezes a derivada da segunda, mais o produto da primeira e da segunda funções vezes a derivada da terceira. Esse resultado pode ser obtido pela aplicação repetida de (7.4). Em primeiro lugar, tratamos o produto $g(x)h(x)$ como uma função única, digamos, $\phi(x)$, de modo que o produto original de três funções torna-se um produto de duas funções, $f(x)\phi(x)$. A isso, podemos aplicar (7.4). Após obter a derivada de $f(x)\phi(x)$, podemos reaplicar (7.4) ao produto $g(x)h(x) \equiv \phi(x)$ para obter $\phi'(x)$. Então, (7.6) será uma consequência. Deixamos os detalhes para você, como exercício.

A validade de uma regra é uma coisa; sua utilidade prática é outra. Por que precisamos da regra do produto quando podemos recorrer ao procedimento alternativo de multiplicar duas funções $f(x)$ e $g(x)$ e então tomar a derivada do produto diretamente? Uma resposta a essa pergunta é que o procedimento alternativo é aplicável apenas a funções *específicas* (numéricas ou paramétricas), ao passo que a regra do produto é aplicável mesmo quando as funções são dadas na forma *geral*. Vamos ilustrar com um exemplo da economia.

Calculando a função de renda marginal a partir da função de renda média

Se tivermos uma dada função de renda média (AR) em forma específica,

$$AR = 15 - Q$$

a função de renda marginal (MR) pode ser achada primeiramente multiplicando AR por Q para obter a função de renda total (R):

$$R \equiv AR \cdot Q = (15 - Q)Q = 15Q - Q^2$$

e, então, diferenciando R :

$$MR \equiv \frac{dR}{dQ} = 15 - 2Q$$

Mas, se a função AR for dada na forma geral $AR = f(Q)$, então a função de renda total também será expressa em uma forma geral:

$$R \equiv AR \cdot Q = f(Q) \cdot Q$$

e, portanto, a abordagem da “multiplicação” será em vão. Entretanto, como R é um produto de duas funções de Q , a saber, $f(Q)$ e a própria Q , a regra do produto pode ser posta em funcionamento. Assim, podemos diferenciar R para obter a função MR como se segue:

$$MR \equiv \frac{dR}{dQ} = f(Q) \cdot 1 + Q \cdot f'(Q) = f(Q) + Qf'(Q) \quad (7.7)$$

Contudo, um resultado geral como esse pode nos dizer algo significativo sobre a MR? Na verdade, pode. Lembrando que $f(Q)$ denota a função AR, vamos rearranjar (7.7) e escrever:

$$MR - AR = MR - f(Q) = Qf'(Q) \quad (7.7')$$

Isso nos dá uma importante relação entre MR e AR: a saber, a diferença entre elas será sempre a quantidade $Qf'(Q)$.

Resta examinar a expressão $Qf'(Q)$. Seu primeiro componente Q denota produção e é sempre não-negativo. O outro componente, $f'(Q)$, representa a inclinação da curva AR em função de Q . Visto que “renda média” e “preço” são apenas nomes diferentes para a mesma coisa:

$$AR \equiv \frac{R}{Q} \equiv \frac{PQ}{Q} \equiv P$$

a curva AR também pode ser considerada como uma curva que relaciona o preço P com a produção Q : $P = f(Q)$. Vista sob essa perspectiva, a curva AR é simplesmente o *inverso* da curva de demanda para o produto da empresa, isto é, as curvas de demanda e a curva AR traçadas nos eixos P e Q são inversas. Sob concorrência pura, a curva AR é uma reta horizontal, de modo que $f'(Q) = 0$ e, conforme (7.7'), $MR - AR = 0$ para todos os valores possíveis de Q . Assim, a curva MR e a curva AR devem coincidir. Sob concorrência imperfeita, por outro lado, a inclinação da curva AR é normalmente decrescente, como ilustra a Figura 7.2, de modo que $f'(Q) < 0$ e, conforme (7.7'), $MR - AR < 0$ para todos os níveis positivos de produção. Nesse caso, a curva MR deve ficar abaixo da curva AR.

A conclusão que acabamos de apresentar é de natureza *qualitativa*; refere-se apenas às posições relativas das duas curvas. Mas (7.7') também fornece a informação *quantitativa* que nos diz que, para qualquer nível de produção Q , a curva MR ficará abaixo da curva AR pela exata quantidade $Q f'(Q)$. Vamos examinar novamente a Figura 7.2 e considerar o nível específico de produção, N . Para essa produção, a expressão $Q f'(Q)$ torna-se, especificamente, $N f'(N)$; se pudermos achar a grandeza de $N f'(N)$ no diagrama, saberemos quanto mais abaixo do ponto G da receita média deve estar o ponto correspondente da receita marginal.

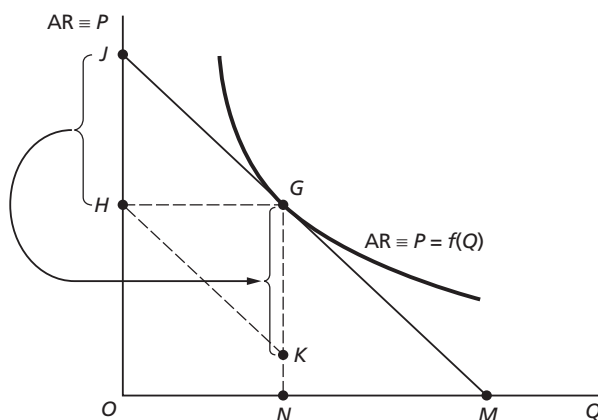
A magnitude de N já está especificada. E $f'(N)$ é simplesmente a inclinação da curva AR no ponto G (onde $Q = N$), isto é, a inclinação da reta tangente JM medida pela razão entre duas distâncias, OJ/OM . Contudo, vemos que $OJ/OM = HJ/HG$; além disso, a distância HG é exatamente a quantidade de produção sob consideração, N . Assim, a distância $N f'(N)$ à qual a curva MR deve estar, abaixo da curva AR na produção N , é

$$N f'(N) = HG \frac{HJ}{HG} = HJ$$

Portanto, se marcarmos uma distância vertical $KG = HJ$ diretamente abaixo do ponto G , então o ponto K deve ser um ponto na curva MR. (Um modo simples de traçar KG com exatidão é desenhar uma linha reta passando pelo ponto H e paralela a JG ; o ponto K é o ponto em que a reta intercepta a reta vertical NG .)

O mesmo procedimento pode ser usado para localizar outros pontos na curva MR. Basta que, para qualquer ponto G' escolhido na curva, primeiramente desenhemos uma tangente à curva AR em G' que interceptará o eixo vertical em algum ponto J' . Então, desenhamos uma reta horizontal de G' até o eixo vertical e denominamos a interseção com o eixo H' . Se marcarmos uma distância vertical $K'G' = H'J'$ diretamente abaixo do ponto G' , então o ponto K' será um ponto sobre a curva MR. Este é o modo gráfico de derivar uma curva MR a partir de uma curva AR dada. Em termos estritos, o desenho preciso de uma reta tangente requer o conhecimento do valor da derivada na produção relevante, isto é, $f'(N)$; por conseguinte, o método gráfico que acabamos de apresentar não pode existir muito bem por si só. Uma exceção importante é o caso

FIGURA 7.2



de uma curva linear AR, na qual a tangente em qualquer ponto da curva é simplesmente a própria reta dada, de modo que, na verdade, não há necessidade de desenhar absolutamente nenhuma tangente. Então, o modo gráfico pode ser aplicado diretamente.

Regra do quociente

A derivada do quociente de duas funções, $f(x)/g(x)$, é:

$$\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

No numerador da expressão do lado direito, achamos dois termos de produto, cada um envolvendo a derivada de somente uma das duas funções originais. Note que $f'(x)$ aparece no termo positivo, e $g'(x)$ no termo negativo. O denominador consiste no quadrado da função $g(x)$; isto é, $g^2(x) \equiv [g(x)]^2$.

EXEMPLO 6

$$\frac{d}{dx} \left(\frac{2x-3}{x+1} \right) = \frac{2(x+1) - (2x-3)(1)}{(x+1)^2} = \frac{5}{(x+1)^2}$$

EXEMPLO 7

$$\frac{d}{dx} \left(\frac{5x}{x^2+1} \right) = \frac{5(x^2+1) - 5x(2x)}{(x^2+1)^2} = \frac{5(1-x^2)}{(x^2+1)^2}$$

EXEMPLO 8

$$\begin{aligned} \frac{d}{dx} \left(\frac{ax^2+b}{cx} \right) &= \frac{2ax(cx) - (ax^2+b)(c)}{(cx)^2} \\ &= \frac{c(ax^2-b)}{(cx)^2} = \frac{ax^2-b}{cx^2} \end{aligned}$$

Essa regra pode ser demonstrada da seguinte maneira. Para qualquer valor de $x = N$, temos:

$$\left. \frac{d}{dx} \frac{f(x)}{g(x)} \right|_{x=N} = \lim_{x \rightarrow N} \frac{f(x)/g(x) - f(N)/g(N)}{x - N} \quad (7.8)$$

A expressão do quociente em seguida ao sinal de limite pode ser reescrita na forma:

$$\frac{f(x)g(N) - f(N)g(x)}{g(x)g(N)} \cdot \frac{1}{x - N}$$

Somando e subtraindo $f(N)g(N)$ no numerador e rearranjando, podemos transformar a expressão para

$$\begin{aligned} &\frac{1}{g(x)g(N)} \left[\frac{f(x)g(N) - f(N)g(N) + f(N)g(N) - f(N)g(x)}{x - N} \right] \\ &= \frac{1}{g(x)g(N)} \left[g(N) \frac{f(x) - f(N)}{x - N} - f(N) \frac{g(x) - g(N)}{x - N} \right] \end{aligned}$$

Substituindo esse resultado em (7.8) e tomando o limite, temos:

$$\left. \frac{d}{dx} \frac{f(x)}{g(x)} \right|_{x=N} = \lim_{x \rightarrow N} \frac{1}{g(x)g(N)} \left[\lim_{x \rightarrow N} g(N) \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N} \right]$$

$$\begin{aligned}
 & - \lim_{x \rightarrow N} f(N) \lim_{x \rightarrow N} \frac{g(x) - g(N)}{x - N} \Big] \\
 & = \frac{1}{g^2(N)} [g(N)f'(N) - f(N)g'(N)] \quad [\text{por (6.13)}]
 \end{aligned}$$

que pode ser generalizada trocando o símbolo N por x , porque N representa qualquer valor de x . Isso prova a regra do quociente.

Relação entre funções de custo marginal e de custo médio

Como uma aplicação da regra do quociente na economia, vamos considerar a taxa de variação do custo médio com a variação da produção.

Dada uma função de custo total $C = C(Q)$, a função de custo médio (AC) é um quociente de duas funções de Q , já que $AC \equiv C(Q)/Q$, definida para $Q > 0$. Por conseguinte, a taxa de variação de AC em relação a Q pode ser achada diferenciando AC:

$$\frac{d}{dQ} \frac{C(Q)}{Q} = \frac{[C'(Q) \cdot Q - C(Q) \cdot 1]}{Q^2} = \frac{1}{Q} \left[C'(Q) - \frac{C(Q)}{Q} \right] \quad (7.9)$$

Disso, segue-se que, para $Q > 0$,

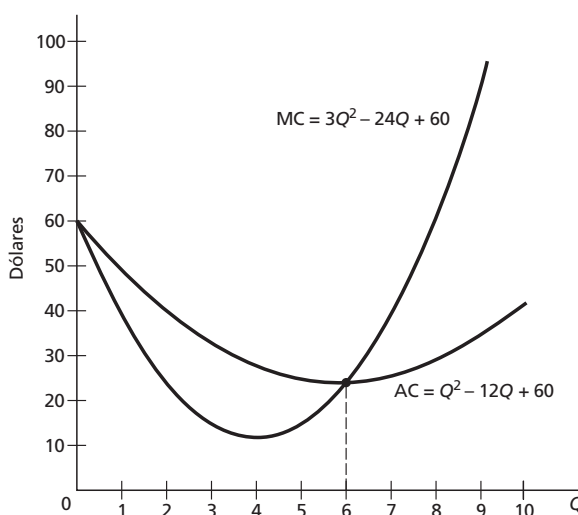
$$\frac{d}{dQ} \frac{C(Q)}{Q} \begin{matrix} \geq \\ \leq \end{matrix} 0 \quad \text{se} \quad C'(Q) \begin{matrix} \geq \\ \leq \end{matrix} \frac{C(Q)}{Q} \quad (7.10)$$

Visto que a derivada $C'(Q)$ representa a função de custo marginal (MC), e $C(Q)/Q$ representa a função AC, o significado econômico de (7.10) é: a inclinação da curva AC será positiva, zero ou negativa se e somente se a curva de custo marginal ficar acima, interceptar ou ficar abaixo da curva AC. Isso é ilustrado na Figura 7.3, onde as funções MC e AC aí traçadas são baseadas na função de custo total específica:

$$C = Q^3 - 12Q^2 + 60Q$$

À esquerda de $Q = 6$, AC é decrescente e, assim, MC fica abaixo dela; à direita, o oposto é verdadeiro. Em $Q = 6$, AC tem inclinação zero, e MC e AC têm o mesmo valor.[†]

FIGURA 7.3



[†] Note que (7.10) não afirma que, quando a inclinação de AC for negativa, a inclinação de MC também tem de ser negativa; ela apenas afirma que AC deve ser maior que MC naquela circunstância. Em $Q = 5$ na Figura 7.3, por exemplo, AC é decrescente, mas MC é ascendente, de modo que suas inclinações terão sinais opostos.

A conclusão qualitativa em (7.10) é enunciada explicitamente em termos de funções de custo. Contudo, sua validade não é afetada se interpretarmos $C(Q)$ como *qualquer outra* função total diferenciável cujas funções média e marginal correspondentes são $C(Q)/Q$ e $C'(Q)$. Assim, esse resultado nos dá uma relação *geral* entre marginal e médio. Em particular, podemos salientar que o fato de MR ficar abaixo de AR quando a inclinação de AR for descendente, como discutido em relação à Figura 7.2, nada mais é do que um caso especial do resultado geral em (7.10).

EXERCÍCIO 7.2

- Dada a função de custo total $C = Q^3 - 5Q^2 + 12Q + 75$, escreva uma função de custo variável (VC). Encontre a derivada da função VC e interprete o significado econômico daquela derivada.
- Dada a função de custo médio $AC = Q^2 - 4Q + 174$, encontre a função MC. A função dada é mais apropriada como função de longo prazo ou de curto prazo? Por quê?
- Diferencie as seguintes expressões utilizando a regra do produto:

(a) $(9x^2 - 2)(3x + 1)$	(c) $x^2(4x + 6)$	(e) $(2 - 3x)(1 + x)(x + 2)$
(b) $(3x + 10)(6x^2 - 7x)$	(d) $(ax - b)(cx^2)$	(f) $(x^2 + 3)x^{-1}$
- (a) Dada $AR = 60 - 3Q$, trace a curva de renda média e depois encontre a curva MR pelo método utilizado na Figura 7.2.
 (b) Encontre a função de renda total e a função de renda marginal por métodos matemáticos, a partir da função AR dada.
 (c) A curva MR derivada por meios gráficos em (a) é compatível com a função MR diferenciada por métodos matemáticos em (b)?
 (d) Comparando as funções AR e MR, o que você pode concluir sobre suas inclinações relativas?
- Dê uma prova matemática para o seguinte resultado geral: dada uma curva média *linear*, a curva marginal correspondente deve ter o mesmo ponto de interseção vertical, mas será duas vezes mais inclinada que a curva média.
- Demonstre o resultado em (7.6) primeiro tratando $g(x)h(x)$ como uma função única, $g(x)h(x) \equiv \phi(x)$ e, depois, aplicando a regra do produto (7.4).
- Calcule as derivadas de:

(a) $(x^2 + 3)/x$	(c) $6x/(x + 5)$
(b) $(x + 9)/x$	(d) $(ax^2 + b)/(cx + d)$
- Dada a função $f(x) = ax + b$, calcule as derivadas de:

(a) $f(x)$	(b) $x f(x)$	(c) $1/f(x)$	(d) $f(x)/x$
------------	--------------	--------------	--------------
- (a) É verdade que $f \in C' \Rightarrow f \in C$?
 (b) É verdade que $f \in C \Rightarrow f \in C'$?
- Calcule as funções marginal e média para as seguintes funções totais e represente os resultados em gráfico.

Função de custo total:

 (a) $C = 3Q^2 + 7Q + 12$
 Função de renda total:
 (b) $R = 10Q - Q^2$
 Função de produto total:
 (c) $Q = aL + bL^2 - cL^3$ ($a, b, c > 0$)

7.3 Regras de diferenciação envolvendo funções de variáveis diferentes

Na Seção 7.2, discutimos as regras de diferenciação de uma soma, diferença, produto ou quociente de duas (ou mais) funções diferenciáveis da mesma variável. Agora, consideraremos casos em que há duas ou mais funções diferenciáveis, cada qual com uma variável independente *distinta*.

Regra da cadeia

Se tivermos uma função diferenciável $z = f(y)$, onde y é, por sua vez, uma função diferenciável de uma outra variável x , digamos, $y = g(x)$, então a derivada de z em relação a x é igual à derivada de z em relação a y , vezes a derivada de y em relação a x . Expressa simbolicamente,

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = f'(y)g'(x) \quad (7.11)$$

Essa regra, conhecida como a *regra da cadeia*, é muito intuitiva. Dado um Δx , deve resultar um Δy correspondente via a função $y = g(x)$, mas essa Δy , por sua vez, resultará em um Δz via a função $z = f(y)$. Assim, há uma “reação em cadeia”, como se segue:

$$\Delta x \xrightarrow{\text{via } g} \Delta y \xrightarrow{\text{via } f} \Delta z$$

Os dois elos dessa cadeia acarretam dois quocientes de diferenças, $\Delta y/\Delta x$ e $\Delta z/\Delta y$, mas, quando esses quocientes são multiplicados, Δy será cancelado e ficaremos com

$$\frac{\Delta z}{\Delta y} \frac{\Delta y}{\Delta x} = \frac{\Delta z}{\Delta x}$$

um quociente de diferenças diferente que relaciona Δz com Δx . Se tomarmos o limite desses quocientes de diferenças quando $\Delta x \rightarrow 0$ (o que implica $\Delta y \rightarrow 0$), cada quociente de diferenças, por sua vez, se transformará em uma derivada; isto é, teremos $(dz/dy)(dy/dx) = dz/dx$. Esse é exatamente o resultado em (7.11).

Em vista da função $y = g(x)$, podemos expressar a função $z = f(y)$ como $z = f[g(x)]$, onde o aparecimento contíguo dos dois símbolos de funções f e g indica que essa é uma *função composta* (função de uma função). É por essa razão que a regra da cadeia também é denominada *regra da função composta* ou *regra da função de uma função*.

A extensão da regra da cadeia a três ou mais funções é direta. Se tivermos $z = f(y)$, $y = g(x)$ e $x = h(w)$, então:

$$\frac{dz}{dw} = \frac{dz}{dy} \frac{dy}{dx} \frac{dx}{dw} = f'(y)g'(x)h'(w)$$

e da mesma forma para casos em que estão envolvidas mais funções.

EXEMPLO 1 Se $z = 3y^2$, onde $y = 2x + 5$, então

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = 6y(2) = 12y = 12(2x + 5)$$

EXEMPLO 2 Se $z = y - 3$, onde $y = x^3$, então

$$\frac{dz}{dx} = 1(3x^2) = 3x^2$$

EXEMPLO 3 A utilidade dessa regra pode ser melhor avaliada quando temos de diferenciar uma função como $z = (x^2 + 3x - 2)^{17}$. Sem a regra da cadeia à nossa disposição, dz/dx somente poderia ser achado por um caminho trabalhoso, isto é, primeiramente elevando a expressão à potência 17. Com a regra da cadeia, contudo, podemos tomar um atalho definindo uma nova variável *intermediária* $y = x^2 + 3x - 2$, de modo a obter, com efeito, duas funções ligadas em cadeia:

$$z = y^{17} \quad \text{e} \quad y = x^2 + 3x - 2$$

Então, a derivada dz/dx pode ser calculada como se segue:

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = 17y^{16}(2x+3) = 17(x^2+3x-2)^{16}(2x+3)$$

EXEMPLO 4

Dada uma função de receita total de uma empresa, $R = f(Q)$, onde a produção Q é uma função do insumo mão-de-obra L , ou $Q = g(L)$, calcule dR/dL . Pela regra da cadeia, temos:

$$\frac{dR}{dL} = \frac{dR}{dQ} \frac{dQ}{dL} = f'(Q)g'(L)$$

Traduzindo para termos econômicos, dR/dQ é a função MR e dQ/dL é a função do produto marginal da mão-de-obra física (MPP_L). De modo semelhante, dR/dL tem a conotação da função do produto marginal da receita da mão-de-obra (MRP_L). Assim, o resultado mostrado constitui o enunciado matemático do seguinte resultado bem conhecido na economia: $MRP_L = MR \cdot MPP_L$.

Regra da função inversa

Se a função $y = f(x)$ representar um mapeamento um a um, isto é, a função é tal que cada valor de y é associado com um único valor de x , a função f terá uma *função inversa* $x = f^{-1}(y)$ (lê-se “ x é uma função inversa de y ”). Aqui, o símbolo f^{-1} é um símbolo de função, que, do mesmo modo que o símbolo de função derivada f' , significa uma função relacionada à função f ; não significa a recíproca da função $f(x)$.

A existência de uma função inversa significa essencialmente que, nesse caso, não somente um dado valor x resultará em um único valor de y [isto é, $y = f(x)$], mas também que um dado valor de y resultará em um único valor de x . Considerando um exemplo não-numérico, podemos exemplificar o mapeamento um a um como o mapeamento do conjunto de todos os maridos para o conjunto de todas as esposas em uma sociedade monógama. Cada marido tem uma única esposa e cada esposa tem um único marido. Ao contrário, o mapeamento do conjunto de todos os pais para o conjunto de todos os filhos não é um a um, porque um pai pode ter mais de um filho, embora cada filho tenha apenas um único pai.

Quando x e y se referem especificamente a números, a propriedade do mapeamento um a um é considerada exclusiva à classe de funções conhecida como *funções estritamente monotônicas* (ou *monótonas*). Dada uma função $f(x)$, se valores sucessivamente maiores da variável independente x *sempre* resultarem em valores sucessivamente maiores de $f(x)$, isto é, se

$$x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$$

então a função f é denominada uma função *estritamente crescente*. Se, por outro lado, aumentos sucessivos em x *sempre* resultarem em reduções sucessivas em $f(x)$, isto é, se

$$x_1 > x_2 \Rightarrow f(x_1) < f(x_2)$$

então diz-se que a função é *estritamente decrescente*. Em qualquer desses casos, a função inversa f^{-1} existe.[†]

Um modo prático para certificar o caráter monotônico (monotonicidade) de dada função $y = f(x)$ é verificar se a derivada $f'(x)$ conserva sempre o mesmo sinal algébrico (não zero) para

[†] Se omitirmos o advérbio *estritamente*, podemos definir funções *monotônicas* (ou *monótonas*) como se segue: uma *função crescente* é uma função que tem a propriedade de

$$x_1 > x_2 \Rightarrow f(x_1) \geq f(x_2) \quad [\text{com a desigualdade fraca } \geq]$$

e uma *função decrescente* é uma função que tem a propriedade de

$$x_1 > x_2 \Rightarrow f(x_1) \leq f(x_2) \quad [\text{com a desigualdade fraca } \leq]$$

Note que, segundo essa definição, uma função escalonada ascendente (descendente) qualifica-se como uma função crescente (decrescente) a despeito do fato de seu gráfico conter segmentos horizontais. Visto que tais funções não têm um mapeamento um a um, não têm funções inversas.

todos os valores de x . Em termos geométricos, isso significa que sua inclinação é sempre ascendente ou é sempre descendente. Assim, a curva de demanda de uma empresa $Q = f(P)$ cuja inclinação é negativa em toda a sua extensão é estritamente decrescente. Como tal, tem uma função inversa $P = f^{-1}(Q)$, que, como já mencionamos, dá a curva de receita média da empresa, já que $P \equiv AR$.

EXEMPLO 5 A função

$$y = 5x + 25$$

tem a derivada $dy/dx = 5$, que é positiva independentemente do valor de x ; assim, a função é estritamente crescente. Segue que existe uma função inversa. No caso presente, a função inversa é calculada com facilidade resolvendo-se a equação dada $y = 5x + 25$ for x . O resultado é a função

$$x = \frac{1}{5}y - 5$$

É interessante notar que essa função inversa também é estritamente crescente, porque $dx/dy = \frac{1}{5} > 0$ para todos os valores de y .

Em termos gerais, se existir uma função inversa, a função original e a função inversa devem ser, ambas, estritamente monotônicas. Além disso, se f^{-1} é a função inversa de f , então f deve ser a função inversa de f^{-1} ; isto é, f e f^{-1} devem ser funções inversas uma da outra.

É fácil verificar que os gráficos de $y = f(x)$ e $x = f^{-1}(y)$ são exatamente os mesmos, apenas com os eixos invertidos. Se colocarmos o eixo x do gráfico de f^{-1} sobre o eixo x do gráfico de f (e fizermos o mesmo para o eixo y), as duas curvas coincidirão. Por outro lado, se o eixo x do gráfico de f^{-1} for sobreposto ao eixo y do gráfico de f (e vice-versa), as duas curvas se tornarão *imagens especulares* uma da outra em relação à reta a 45° que passa pela origem. Essa relação de imagem especular nos proporciona um modo fácil de representar graficamente a função inversa f^{-1} , uma vez dado o gráfico da função original f . (Você deve tentar fazer isso com as duas funções no Exemplo 5.)

Para funções inversas, a regra de diferenciação é

$$\frac{dx}{dy} = \frac{1}{dy/dx}$$

Isso significa que a derivada da função inversa é a recíproca da derivada da função original; assim, dx/dy deve adotar o mesmo sinal de dy/dx , de modo que, se f for estritamente crescente (decrescente), então f^{-1} também deverá ser.

Para verificar essa regra, podemos nos referir ao Exemplo 5, onde achamos dy/dx igual a 5 e dx/dy igual a $\frac{1}{5}$. Essas duas derivadas são, de fato, recíprocas uma da outra e têm o mesmo sinal.

Naquele exemplo simples, a função inversa é relativamente fácil de obter, de modo que sua derivada dx/dy pode ser achada diretamente a partir da função inversa. Entretanto, como mostra o Exemplo 6, às vezes é difícil expressar a função inversa explicitamente e, assim, a diferenciação direta pode não ser viável. Então, a utilidade da regra da função inversa fica bem mais aparente.

EXEMPLO 6 Dado $y = x^5 + x$, ache dx/dy . Antes de mais nada, visto que

$$\frac{dy}{dx} = 5x^4 + 1 > 0$$

para qualquer valor de x , a função dada é estritamente crescente e existe uma função inversa. Resolver a equação dada para x pode não ser uma tarefa assim tão fácil, porém, não obstante, a derivada da função inversa pode ser calculada rapidamente utilizando-se a regra da função inversa:

$$\frac{dx}{dy} = \frac{1}{dy/dx} = \frac{1}{5x^4 + 1}$$

A regra da função inversa, em termos estritos, é aplicável somente quando a função envolvida é um mapeamento um a um. Contudo, na realidade, temos algum espaço de manobra. Por exemplo, quando se tratar de uma curva em forma de U (não estritamente monotônica), podemos considerar os segmentos de inclinação descendente e ascendente como duas funções *separadas*, cada uma com um domínio restrito e cada uma estritamente monotônica no domínio restrito. Então, a regra da função inversa pode ser novamente aplicada a cada uma delas.

EXERCÍCIO 7.3

1. Dado $y = u^3 + 2u$, onde $u = 5 - x^2$, calcule dy/dx pela regra da cadeia.
2. Dados $w = ay^2$ e $y = bx^2 + cx$, calcule dw/dx pela regra da cadeia.
3. Use a regra da cadeia para calcular dy/dx para as seguintes:
 - (a) $y = (3x^2 - 13)^3$
 - (b) $y = (7x^3 - 5)^9$
 - (c) $y = (ax + b)^5$
4. Dado $y = (16x + 3)^{-2}$, use a regra da cadeia para calcular dy/dx . Depois, reescreva a função como $y = 1/(16x + 3)^2$ e calcule dy/dx pela regra do quociente. As respostas são idênticas?
5. Dado $y = 7x + 21$, calcule sua função inversa. Depois, calcule dy/dx e dx/dy , e verifique a regra da função inversa. Verifique também se os gráficos das duas funções apresentam a relação da imagem especular entre elas.
6. As funções seguintes são estritamente monotônicas?
 - (a) $y = -x^6 + 5 \quad (x > 0)$
 - (b) $y = 4x^5 + x^3 + 3x$
 Para cada função estritamente monotônica, calcule dx/dy pela regra da função inversa.

7.4 Diferenciação parcial

Até aqui, consideramos somente as derivadas de funções de uma única variável independente. Contudo, na análise estática comparativa, provavelmente encontraremos a situação na qual aparecem vários parâmetros em um modelo, de modo que o valor de equilíbrio de cada variável endógena pode ser uma função de mais de um parâmetro. Portanto, como preparação final para a aplicação do conceito de derivada à estática comparativa, devemos aprender como calcular a derivada de uma função de mais de uma variável.

Derivadas parciais

Vamos considerar uma função

$$y = f(x_1, x_2, \dots, x_n) \quad (7.12)$$

na qual as variáveis x_i ($i = 1, 2, \dots, n$) são todas *independentes* umas das outras, de modo que cada uma delas pode variar por si só sem afetar as outras. Se a variável x_1 sofrer uma variação Δx_1 enquanto $x_2, \dots, x_n$ permanecem fixas, haverá uma variação correspondente em y , a saber, Δy . O quociente de diferenças, nesse caso, pode ser expresso como:

$$\frac{\Delta y}{\Delta x_1} = \frac{f(x_1 + \Delta x_1, x_2, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{\Delta x_1} \quad (7.13)$$

Se tomarmos o limite de $\Delta y/\Delta x_1$ quando $\Delta x_1 \rightarrow 0$, esse limite constituirá uma derivada, a qual denominamos *derivada parcial de y em relação a x_1* , para indicar que todas as outras variáveis independentes na função são mantidas constantes quando tomamos essa derivada em particular. Derivadas parciais semelhantes podem ser definidas por variações infinitesimais nas outras variáveis independentes. O processo de tomar derivadas parciais é denominado *diferenciação parcial*.

Derivadas parciais recebem símbolos distintivos. Em lugar da letra d (como em dy/dx), empregamos o símbolo ∂ , que é uma variante da letra grega δ (delta minúscula). Assim, daqui em diante, escreveremos $\partial y/\partial x_i$ que se lê: “a derivada parcial de y em relação a x_i ”. O símbolo de derivada parcial às vezes também é escrito como $\frac{\partial}{\partial x_i} y$; nesse caso, a parte $\partial/\partial x_i$ pode ser conside-

rada como um símbolo operador que indica que devemos tomar a derivada parcial de (alguma função) com relação à variável x_i . Visto que a função envolvida aqui é denotada por f em (7.12), também podemos escrever $\partial f/\partial x_i$.

No caso da derivada parcial, também há uma contraparte para o símbolo $f'(x)$ que utilizamos antes? A resposta é sim. Em vez de f' , entretanto, agora usamos f_1, f_2 etc., onde o índice mostra qual variável independente (isolada) estará variando. Se a função em (7.12) por acaso for escrita em termos de variáveis sem índices, tal como $y = f(u, v, w)$, então as derivadas parciais podem ser denotadas por f_u, f_v e f_w em vez de f_1, f_2 e f_3 .

Conforme essas notações e tomando como base (7.12) e (7.13), agora podemos definir:

$$f_1 \equiv \frac{\partial y}{\partial x_1} \equiv \lim_{\Delta x_1 \rightarrow 0} \frac{\Delta y}{\Delta x_1}$$

como a primeira no conjunto de n derivadas parciais da função f .

Técnicas de diferenciação parcial

A diferenciação parcial é diferente da diferenciação discutida em capítulo anterior, primordialmente porque temos de manter *constantes* ($n - 1$) variáveis independentes enquanto permitimos que *somente uma* varie. Considerando que aprendemos como manipular *constantes* na diferenciação, a diferenciação propriamente dita não deve causar muitos problemas.

EXEMPLO 1

Dado $y = f(x_1, x_2) = 3x_1^2 + x_1x_2 + 4x_2^2$, calcule as derivadas parciais. Para achar $\partial y/\partial x_1$, (ou f_1), temos de ter sempre em mente que x_2 deve ser tratada como uma constante durante a diferenciação. Como tal, x_2 será eliminada no processo se for uma constante *aditiva* (tal como no termo $4x_2^2$), mas será conservada se for uma constante *multiplicativa* (tal como no termo x_1x_2). Assim, temos

$$\frac{\partial y}{\partial x_1} \equiv f_1 = 6x_1 + x_2$$

De modo semelhante, tratando x_1 como uma constante, achamos que

$$\frac{\partial y}{\partial x_2} \equiv f_2 = x_1 + 8x_2$$

Note que, assim como a função primitiva f , ambas as derivadas parciais são, elas mesmas, funções das variáveis x_1 e x_2 . Isto é, podemos escrevê-las como duas funções derivadas

$$f_1 = f_1(x_1, x_2) \quad \text{e} \quad f_2 = f_2(x_1, x_2)$$

Para o ponto $(x_1, x_2) = (1, 3)$ no domínio da função f , por exemplo, as derivadas parciais assumirão os seguintes valores específicos:

$$f_1(1, 3) = 6(1) + 3 = 9 \quad \text{e} \quad f_2(1, 3) = 1 + 8(3) = 25$$

EXEMPLO 2

Dado $y = f(u, v) = (u + 4)(3u + 2v)$, as derivadas parciais podem ser calculadas utilizando-se a regra do produto. Mantendo v constante, temos

$$f_u = (u + 4)(3) + 1(3u + 2v) = 2(3u + v + 6)$$

De modo semelhante, mantendo u constante, encontramos

$$f_v = (u + 4)(2) + 0(3u + 2v) = 2(u + 4)$$

Quando $u = 2$ e $v = 1$, essas derivadas assumirão os seguintes valores:

$$f_u(2, 1) = 2(13) = 26 \quad \text{e} \quad f_v(2, 1) = 2(6) = 12$$

EXEMPLO 3

Dado $y = (3u - 2v)/(u^2 + 3v)$, as derivadas parciais podem ser calculadas utilizando-se a regra do quociente:

$$\frac{\partial y}{\partial u} = \frac{3(u^2 + 3v) - 2u(3u - 2v)}{(u^2 + 3v)^2} = \frac{-3u^2 + 4uv + 9v}{(u^2 + 3v)^2}$$

$$\frac{\partial y}{\partial v} = \frac{-2(u^2 + 3v) - 3(3u - 2v)}{(u^2 + 3v)^2} = \frac{-u(2u + 9)}{(u^2 + 3v)^2}$$

Interpretação geométrica de derivadas parciais

Sendo um tipo especial de derivada, uma derivada parcial é uma medição das taxas instantâneas de variação de alguma variável e, como tal, novamente tem uma contraparte geométrica na inclinação de uma curva específica.

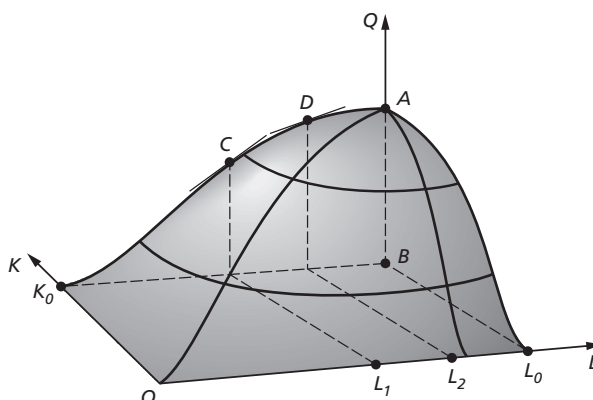
Vamos considerar a função de produção $Q = Q(K, L)$, onde Q , K e L denotam produção, insumo de capital e insumo de mão-de-obra, respectivamente. Essa função é uma versão particular de duas variáveis de (7.12), com $n = 2$. Portanto, podemos definir duas derivadas parciais $\partial Q / \partial K$ (ou Q_K) e $\partial Q / \partial L$ (ou Q_L). A derivada parcial Q_K é relativa às taxas de variação da produção segundo variações infinitesimais no capital, enquanto o insumo de mão-de-obra é mantido constante. Assim, Q_K simboliza a função do produto marginal do capital físico (MPP_K). De modo semelhante, a derivada parcial Q_L é a representação matemática da função MPP_L.

Em termos geométricos, a função de produção $Q = Q(K, L)$ pode ser representada por uma *superfície de produção* em um espaço tridimensional, como mostra a Figura 7.4. A variável Q é traçada na vertical, de modo que, para qualquer ponto (K, L) no plano de base (plano KL), a altura da superfície indicará a produção Q . O domínio da função deve consistir em todo o quadrante não-negativo do plano base, mas, para nossa finalidade, basta considerar um subconjunto do domínio: o retângulo OK_0BL_0 . Como consequência, somente uma pequena porção da superfície é mostrada na figura.

Agora vamos manter o capital fixo no nível K_0 e considerar apenas variações no insumo L . Estabelecendo $K = K_0$, todos os pontos em nosso domínio (abreviado) tornam-se irrelevantes, exceto os que se encontram sobre o segmento de reta K_0B . Pelo mesmo critério, somente a curva K_0CDA (uma seção transversal da superfície de produção) é pertinente à presente discussão. Essa curva representa uma curva de produto total do trabalho físico (TPP_L) para uma quantidade fixa de capital $K = K_0$; assim, podemos ler, pela sua inclinação, a taxa de variação de Q relativa às variações em L quando K é mantido constante. Portanto, fica claro que a inclinação de uma curva como K_0CDA representa a contraparte geométrica da derivada parcial Q_L . Mais uma vez, notamos que a inclinação de uma curva total (TPP_L) é sua curva marginal correspondente (MPP_L $\equiv$ Q_L).

Como mencionamos antes, uma derivada parcial é uma função de todas as variáveis independentes da função primitiva. O fato de Q_L ser uma função de L fica imediatamente óbvio pela própria curva K_0CDA . Quando $L = L_1$, o valor de Q_L é igual à inclinação da curva no ponto C ; mas, quando $L = L_2$, a inclinação relevante é aquela obtida no ponto D . Por que Q_L também é uma função de K ? A resposta é que K pode ser fixada em vários níveis e, para cada nível fixado de K , resulta uma curva TPP_L diferente (uma seção transversal diferente da superfície de produção), com repercussões inevitáveis sobre a derivada Q_L . Por conseguinte, Q_L também é uma função de K .

Pode-se dar uma interpretação análoga à derivada parcial Q_K . Se, em vez de K , o insumo de mão-de-obra for mantido constante (digamos, no nível de L_0), o segmento de reta L_0B será o subconjunto relevante do domínio e a curva L_0A indicará o subconjunto relevante da superfície de produção. A derivada parcial Q_K , então, pode ser interpretada como a inclinação da curva L_0A – mantendo-se em mente que o eixo K se estende da esquerda para a direita (sudeste para noroeste) na Figura 7.4. Deve-se observar, também, que Q_K é, novamente, uma função de ambas as variáveis L e K .



Todas as derivadas parciais de uma função $y = f(x_1, x_2, \dots, x_n)$ podem ser reunidas sob uma única entidade matemática denominada o *vetor gradiente* ou, simplesmente, o *gradiente* da função f :

onde $f_i \equiv \partial y / \partial x_i$. Note que estamos usando parênteses em vez de chaves quando escrevemos o vetor. O vetor também pode ser denotado por $\nabla f(x_1, x_2, \dots, x_n)$, onde ∇ (lê-se: “del”) é a versão invertida da letra grega Δ .

Visto que a função f tem n argumentos, há, no total, n derivadas parciais; assim, $\text{grad } f$ é um vetor n . Quando essas derivadas são avaliadas em um ponto específico $(x_{10}, x_{20}, \dots, x_{n0})$ no domínio, obtemos $\text{grad } f(x_{10}, x_{20}, \dots, x_{n0})$, um vetor de *valores* específicos de derivadas.

$$\nabla Q = \nabla Q(K, L) = (Q_K, Q_L)$$

1. Calcule $\partial y/\partial x_1$ e $\partial y/\partial x_2$ para cada uma das seguintes funções:

(b) $y = 7x_1 + 6x_1x_2^2 - 9x_2^3$ (d) $y = (5x_1 + 3)/(x_2 - 2)$

2. Calcule f_x e f_y partindo das seguintes:

$$(b) f(x, y) = (x^2 - 3y)(x - 2) \qquad (d) f(x, y) = \frac{x^2 - 1}{xy}$$

3. Com as respostas do Problema 2, calcule $f_x(1, 2)$ – o valor da derivada parcial f_x quando $x = 1$ e $y = 2$ – para cada função.

4. Dada a função de produção $Q = 96K^{0,3}L^{0,7}$, calcule as funções MPP_K e MPP_L . MPP_K é uma função somente de K , ou de ambos, K e L ? E MPP_L ?

5. A função utilidade de um indivíduo assume a forma

$$U = U(x_1, x_2) = (x_1 + 2)^2(x_2 + 3)^3$$

onde U é utilidade total e x_1 e x_2 são as quantidades de duas mercadorias consumidas:

- (a) Encontre a função de utilidade marginal de cada uma das duas mercadorias.
 - (b) Calcule o valor da utilidade marginal da primeira mercadoria quando são consumidas 3 unidades de cada mercadoria.
6. A oferta total de moeda M tem dois componentes: depósitos bancários, D , e dinheiro vivo, C , entre os quais admitimos existir uma razão constante $C/D = c$, $0 < c < 1$. A moeda de alto poder H é definida como a soma do dinheiro vivo nas mãos do público e as reservas mantidas pelos bancos. Reservas bancárias são uma fração dos depósitos bancários, determinada pela razão de reserva r , $0 < r < 1$.
- (a) Expresse a oferta de moeda M como uma função da moeda de alto poder H .
 - (b) Um aumento na razão de reserva r aumentaria ou reduziria a oferta de moeda?
 - (c) Como um aumento na razão entre dinheiro vivo e depósitos, c , afetaria a oferta de moeda?
7. Escreva os gradientes das seguintes funções:
- (a) $f(x, y, z) = x^2 + y^3 + z^4$
 - (b) $f(x, y, z) = xyz$

7.5 Aplicações à análise estática comparativa

Agora que já conhecemos as várias regras de diferenciação, finalmente podemos atacar o problema proposto pela análise estática comparativa: a saber, qual será a variação do valor de equilíbrio de uma variável endógena quando houver uma variação em qualquer uma das variáveis exógenas ou parâmetros.

Modelo de mercado

Em primeiro lugar, vamos considerar novamente o modelo de mercado simples de uma mercadoria de (3.1). Aquele modelo pode ser escrito na forma de duas equações:

$$Q = a - bP \quad (a, b > 0) \quad [\text{demanda}]$$

$$Q = -c + dP \quad (c, d > 0) \quad [\text{oferta}]$$

com soluções

$$P^* = \frac{a + c}{b + d} \quad (7.14)$$

$$Q^* = \frac{ad - bc}{b + d} \quad (7.15)$$

Diremos que essas soluções estão na *forma reduzida*: as duas variáveis endógenas foram reduzidas a expressões explícitas dos quatro parâmetros mutuamente independentes a , b , c e d .

Para saber como uma variação infinitesimal em um dos parâmetros afetará o valor de P^* , basta diferenciar parcialmente (7.14) em relação a cada um dos parâmetros. Se o *signal* de uma derivada parcial, digamos, $\partial P^* / \partial a$, puder ser determinado pela informação dada sobre os parâmetros, saberemos em que direção o parâmetro P^* se moverá quando o parâmetro a mudar; essa é uma conclusão qualitativa. Se a grandeza de $\partial P^* / \partial a$ puder ser determinada, ela constituirá uma conclusão quantitativa.

De modo semelhante, podemos tirar conclusões qualitativas ou quantitativas das derivadas parciais de Q^* em relação a cada parâmetro, tal como $\partial Q^* / \partial a$. Entretanto, para evitar mal-entendidos, é preciso fazer uma clara distinção entre as duas derivadas $\partial Q^* / \partial a$ e $\partial Q / \partial a$. A última derivada é um conceito adequado à função demanda quando considerada por si só e sem relação

com a função oferta. Por outro lado, a derivada $\partial Q^*/\partial a$ é pertinente à quantidade de equilíbrio em (7.15), a qual, por ter as características de uma solução do modelo, leva em conta a interação conjunta entre demanda e oferta. Para enfatizar essa distinção, nos referiremos às derivadas parciais de P^* e Q^* em relação aos parâmetros como *derivadas estáticas comparativas*. A possibilidade de confusão entre $\partial Q^*/\partial a$ e $\partial Q/\partial a$ é a exata razão por que preferimos utilizar a notação com asterisco, com em Q^* , para denotar o valor de equilíbrio.

Concentrando-nos em P^* , por enquanto, podemos obter, de (7.14), as quatro seguintes derivadas parciais:

$$\frac{\partial P^*}{\partial a} = \frac{1}{b+d} \left[\text{parâmetro } a \text{ tem o coeficiente } \frac{1}{b+d} \right]$$

$$\frac{\partial P^*}{\partial b} = \frac{0(b+d) - 1(a+c)}{(b+d)^2} = \frac{-(a+c)}{(b+d)^2} \quad [\text{regra do quociente}]$$

$$\frac{\partial P^*}{\partial c} = \frac{1}{b+d} \left(= \frac{\partial P^*}{\partial a} \right)$$

$$\frac{\partial P^*}{\partial d} = \frac{0(b+d) - 1(a+c)}{(b+d)^2} = \frac{-(a+c)}{(b+d)^2} \left(= \frac{\partial P^*}{\partial b} \right)$$

Visto que todos os parâmetros estão restritos a serem positivos no presente modelo, podemos concluir que

$$\frac{\partial P^*}{\partial a} = \frac{\partial P^*}{\partial c} > 0 \quad \text{e} \quad \frac{\partial P^*}{\partial b} = \frac{\partial P^*}{\partial d} < 0 \quad (7.16)$$

Para uma apreciação mais completa dos resultados em (7.16), vamos examinar a Figura 7.5, onde cada diagrama mostra uma variação em um dos parâmetros. Como antes, estamos traçando Q (em vez de P) no eixo vertical.

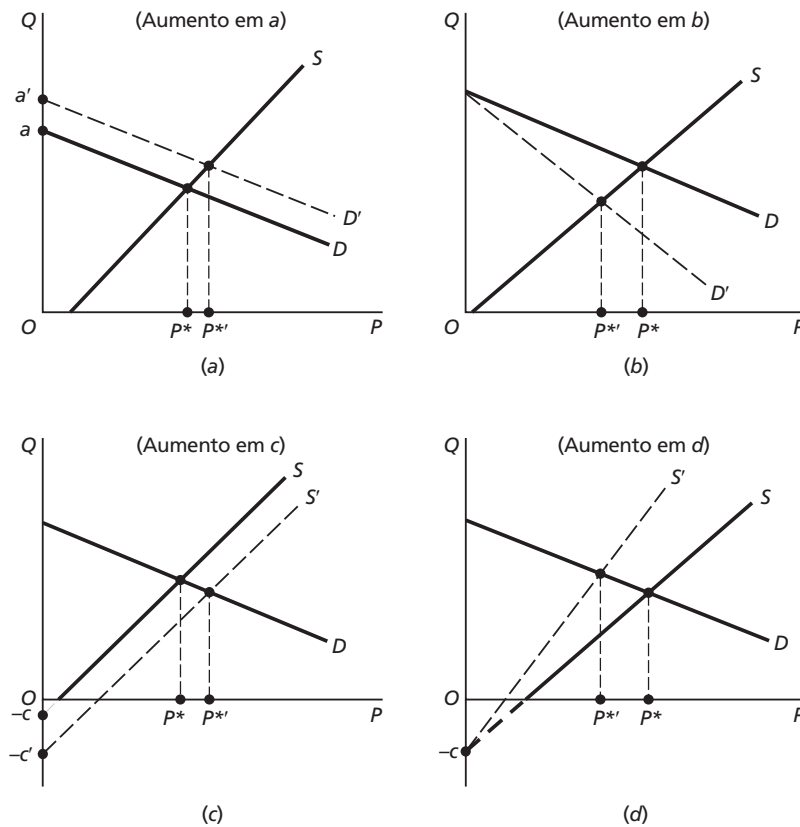
A Figura 7.5a representa um aumento no parâmetro a (para a'). Isso significa uma interseção vertical mais alta para a curva de demanda e, considerando que o parâmetro b (o parâmetro da inclinação) não muda, o aumento resulta em um deslocamento paralelo para cima da curva de demanda de D para D' . A interseção de D' com a curva de oferta S determina um preço de equilíbrio $P^{*'}$, que é maior que o preço de equilíbrio anterior, P^* . Isso corrobora o resultado que $\partial P^*/\partial a > 0$, embora, para facilitar a exposição gráfica, tenhamos mostrado na Figura 7.5a uma variação muito maior no parâmetro a do que o implícito no conceito de derivada.

A interpretação da situação na Figura 7.5c é semelhante; mas, posto que o aumento ocorre no parâmetro c , o resultado é um deslocamento paralelo da curva de oferta. Note que esse deslocamento é para baixo porque a curva de oferta tem uma interseção vertical de $-c$; assim, um aumento em c significaria uma mudança na interseção, digamos, de -2 para -4 . O resultado gráfico da estática comparativa, ou seja, que $P^{*'}$ é maior que P^* , novamente está de acordo com o que o sinal positivo da derivada $\partial P^*/\partial c$ nos levaria a esperar.

As Figuras 7.5b e 7.5d ilustram os efeitos das variações nos parâmetros de inclinação b e d das duas funções no modelo. Um aumento em b significa que a inclinação da curva de demanda assumirá um valor numérico (absoluto) maior, isto é, a curva ficará mais íngreme. De acordo com o resultado $\partial P^*/\partial b < 0$, achamos uma redução em P^* nesse diagrama. O aumento em d que torna a curva de oferta mais acentuada também resulta em uma redução no preço de equilíbrio. Mais uma vez, é claro, isso está de acordo com o sinal negativo da derivada estática comparativa $\partial P^*/\partial d$.

Até aqui, todos os resultados em (7.16) aparentemente foram obtidos por meios gráficos. Se assim for, por que nos incomodariamos de usar a diferenciação? A resposta é que a abordagem da diferenciação apresenta, no mínimo, duas grandes vantagens. Em primeiro lugar, a técnica

FIGURA 7.5



gráfica está sujeita a restrições dimensionais, e a diferenciação, não. Mesmo quando o número de variáveis endógenas e parâmetros for tal que o estado de equilíbrio não pode ser mostrado graficamente, ainda podemos aplicar as técnicas de diferenciação ao problema. Em segundo lugar, o método de diferenciação pode dar resultados com níveis mais altos de generalidade. Os resultados em (7.16) continuarão válidos independentemente dos valores específicos que os parâmetros a , b , c e d assumirem, contanto que eles satisfaçam as restrições de sinal. Portanto, as conclusões da estática comparativa obtidas com esse modelo são, com efeito, aplicáveis a um número infinito de combinações de funções (lineares) de oferta e demanda. Ao contrário, a abordagem gráfica trata somente de alguns membros específicos da família de curvas de oferta e demanda e, em termos estritos, o resultado analítico derivado dessa abordagem é aplicável apenas às funções específicas consideradas.

Essa discussão serve para ilustrar a aplicação da diferenciação parcial à análise estática comparativa do modelo simples de mercado, mas, na verdade, apenas metade da tarefa foi executada, pois também podemos achar as derivadas estáticas comparativas pertinentes a Q^* . Deixamos que você faça isso como exercício.

Modelo de renda nacional

Em vez do modelo simples de renda nacional discutido no Capítulo 3, vamos agora trabalhar com um modelo ligeiramente ampliado, com três variáveis endógenas, Y (renda nacional), C (consumo) e T (impostos):

$$\begin{aligned} Y &= C + I_0 + G_0 \\ C &= \alpha + \beta(Y - T) & (\alpha > 0; \quad 0 < \beta < 1) \\ T &= \gamma + \delta Y & (\gamma > 0; \quad 0 < \delta < 1) \end{aligned} \quad (7.17)$$

A primeira equação desse sistema dá a condição de equilíbrio para a renda nacional, enquanto a segunda e a terceira equações mostram, respectivamente, como C e T são determinados no modelo.

As restrições sobre os valores dos parâmetros α, β, γ , e δ podem ser explicadas assim: α é positivo porque o consumo é positivo mesmo que a renda disponível ($Y - T$) seja zero; β é uma fração positiva porque representa a propensão marginal ao consumo; γ é positivo porque, mesmo que Y seja zero, o governo ainda terá uma receita positiva de impostos (advinda de outras bases de tributação além da renda); e, finalmente, δ é uma fração positiva porque representa uma taxa de imposto sobre a renda e, como tal, não pode passar de 100%. As variáveis exógenas I_0 (investimento) e G_0 (despesas do governo) são, é claro, não-negativas. Admite-se que todos os parâmetros e variáveis exógenas são independentes uns dos outros, de modo que pode-se atribuir um novo valor a qualquer um deles sem afetar os outros.

Esse modelo pode ser resolvido para Y^* substituindo-se a terceira equação de (7.17) na segunda e então substituindo-se a equação resultante na primeira. A renda de equilíbrio (na forma reduzida) é

$$Y^* = \frac{\alpha - \beta\gamma + I_0 + G_0}{1 - \beta + \beta\delta} \quad (7.18)$$

Valores de equilíbrio semelhantes também podem ser achados para as variáveis C e T , mas vamos nos concentrar na renda de equilíbrio.

Pode-se obter seis derivadas estáticas comparativas de (7.18). Entre elas, as três seguintes têm uma significância política especial:

$$\frac{\partial Y^*}{\partial G_0} = \frac{1}{1 - \beta + \beta\delta} > 0 \quad (7.19)$$

$$\frac{\partial Y^*}{\partial \gamma} = \frac{-\beta}{1 - \beta + \beta\delta} < 0 \quad (7.20)$$

$$\frac{\partial Y^*}{\partial \delta} = \frac{-\beta(\alpha - \beta\gamma + I_0 + G_0)}{(1 - \beta + \beta\delta)^2} = \frac{-\beta Y^*}{1 - \beta + \beta\delta} < 0 \quad [\text{por (7.18)}] \quad (7.21)$$

A derivada parcial em (7.19) nos dá o *multiplicador da despesa do governo*. Aqui ele tem sinal positivo porque β é menor que 1, e $\beta\delta$ é maior que zero. Se forem atribuídos valores numéricos aos parâmetros β e δ , podemos também achar o valor numérico desse multiplicador por (7.19). A derivada em (7.20) pode ser denominada *multiplicador tributário não referente à renda* porque mostra como uma variação em γ , a receita do governo advinda de outras fontes que não a renda, afetar a renda de equilíbrio. Esse multiplicador é negativo no presente modelo porque o denominador em (7.20) é positivo e o numerador é negativo. Por fim, a derivada parcial em (7.21) – que não tem as características de um multiplicador, já que não relaciona a variação de um dólar com uma outra variação de um dólar, como fazem as derivadas em (7.19) e (7.20) – nos mostra até que ponto um aumento na taxa do imposto sobre a renda δ reduzirá a renda de equilíbrio.

Note, mais uma vez, a diferença entre as duas derivadas $\partial Y^*/\partial G_0$ e $\partial Y/\partial G_0$. A primeira é derivada de (7.18), a expressão para a renda de equilíbrio. A última, que pode ser obtida da primeira equação em (7.17), é $\partial Y/\partial G_0 = 1$, que é de todo diferente em grandeza e em conceito.

Modelo de insumo-produto

A solução de um modelo aberto de insumo-produto aparece como uma equação matricial $x^* = (I - A)^{-1}d$. Se denotarmos a matriz inversa $(I - A)^{-1}$ por $V = [v_{ij}]$, então, por exemplo, a solução para uma economia de três indústrias pode ser escrita como $x^* = Vd$, ou

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (7.22)$$

Quais são as taxas de variação dos valores de solução x_j^* em relação às demandas exógenas finais d_1 , d_2 e d_3 ? A resposta geral é que

$$\frac{\partial x_j^*}{\partial d_k} = v_{jk} \quad (j, k = 1, 2, 3) \quad (7.23)$$

Para ver isso, vamos multiplicar Vd em (7.22) e expressar a solução como

$$\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11}d_1 + v_{12}d_2 + v_{13}d_3 \\ v_{21}d_1 + v_{22}d_2 + v_{23}d_3 \\ v_{31}d_1 + v_{32}d_2 + v_{33}d_3 \end{bmatrix}$$

Nesse sistema de três equações, cada uma dá um valor de solução específico como uma função das demandas exógenas finais. A diferenciação parcial das três produz um total de nove derivadas estáticas comparativas:

$$\begin{aligned} \frac{\partial x_1^*}{\partial d_1} &= v_{11} & \frac{\partial x_1^*}{\partial d_2} &= v_{12} & \frac{\partial x_1^*}{\partial d_3} &= v_{13} \\ \frac{\partial x_2^*}{\partial d_1} &= v_{21} & \frac{\partial x_2^*}{\partial d_2} &= v_{22} & \frac{\partial x_2^*}{\partial d_3} &= v_{23} \\ \frac{\partial x_3^*}{\partial d_1} &= v_{31} & \frac{\partial x_3^*}{\partial d_2} &= v_{32} & \frac{\partial x_3^*}{\partial d_3} &= v_{33} \end{aligned} \quad (7.23')$$

Essa é simplesmente a versão expandida de (7.23).

Lendo (7.23') como três colunas distintas, podemos combinar as três derivadas em cada coluna em uma derivada de matriz (vetor):

$$\frac{\partial x^*}{\partial d_1} \equiv \frac{\partial}{\partial d_1} \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \end{bmatrix} \quad \frac{\partial x^*}{\partial d_2} = \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \end{bmatrix} \quad \frac{\partial x^*}{\partial d_3} = \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \end{bmatrix} \quad (7.23'')$$

Visto que os três vetores coluna em (7.23'') são meramente as colunas da matriz V , consolidando mais ainda, podemos resumir as nove derivadas em uma única derivada de matriz $\partial x^* / \partial d$. Dado $x^* = Vd$, podemos escrever simplesmente

$$\frac{\partial x^*}{\partial d} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = V \equiv (I - A)^{-1}$$

Assim, $(I - A)^{-1}$, a inversa da matriz de Leontief, nos dá uma exibição ordenada de todas as derivadas estáticas comparativas de nosso modelo aberto de insumo-produto. Obviamente, essa derivada de matriz pode ser estendida com facilidade do modelo de três indústrias para o caso geral de n indústrias.

Derivadas estáticas comparativas do modelo de insumo-produção são ferramentas úteis de planejamento econômico, pois dão a resposta à pergunta: se as metas de planejamento, tais como refletidas em $(d_1, d_2, \dots, d_n)$, forem revisadas, e se quisermos cuidar de todos os requisitos diretos e indiretos na economia, de modo que ela fique completamente livre de gargalos, como devemos mudar as metas de produção das n indústrias?

EXERCÍCIO 7.5

1. Examine as propriedades estáticas comparativas da quantidade de equilíbrio em (7.15), e verifique seus resultados por meio de análise gráfica.
2. Com base em (7.18), calcule as derivadas parciais $\partial Y^* / \partial I_0$, $\partial Y^* / \partial \alpha$ e $\partial Y^* / \partial \beta$. Interprete seus significados e determine seus sinais.
3. O modelo numérico de insumo-produto (5.21) foi resolvido na Seção 5.7.
 - (a) Quantas derivadas estáticas comparativas podem ser derivadas?
 - (b) Escreva essas derivadas na forma de (7.23') e (7.23'').

7.6 Nota sobre determinantes jacobianos

Nosso estudo de derivadas parciais foi motivado exclusivamente por considerações estáticas comparativas. Mas derivadas parciais também proporcionam um meio de testar a existência de dependência funcional (linear ou não-linear) entre um conjunto de n funções de n variáveis. Isso está relacionado à noção de determinantes jacobianos (nome derivado de Jacobi).

Considere as duas funções

$$\begin{aligned} y_1 &= 2x_1 + 3x_2 \\ y_2 &= 4x_1^2 + 12x_1x_2 + 9x_2^2 \end{aligned} \quad (7.24)$$

Se obtivermos todas as quatro derivadas parciais

$$\frac{\partial y_1}{\partial x_1} = 2 \quad \frac{\partial y_1}{\partial x_2} = 3 \quad \frac{\partial y_2}{\partial x_1} = 8x_1 + 12x_2 \quad \frac{\partial y_2}{\partial x_2} = 12x_1 + 18x_2$$

e as arranjarmos em uma matriz quadrada em uma ordem determinada, denominada uma matriz jacobiana e denotada por J , e depois tomarmos seu determinante, o resultado será algo conhecido como um *determinante jacobiano* (ou, abreviadamente, *um jacobiano*), denotado por $|J|$:

$$|J| \equiv \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ (8x_1 + 12x_2) & (12x_1 + 18x_2) \end{vmatrix} \quad (7.25)$$

Para economizar espaço, às vezes esse jacobiano também é expresso como

$$|J| \equiv \left| \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right|$$

Em termos mais gerais, se tivermos n funções diferenciáveis de n variáveis, não necessariamente lineares,

$$y_i = f^i(x_1, \dots, x_n, x_{n+1}, x_{n+2}) \quad (i = 1, 2, \dots, n)$$

Nesse caso, se mantivermos constantes quaisquer duas variáveis (digamos, x_{n+1} e x_{n+2}), ou as tratarmos como parâmetros, teremos novamente n funções de exatamente n variáveis e podemos formar um jacobiano. Além disso, mantendo constante um par diferente das x variáveis, podemos formar um jacobiano diferente. Essa situação será encontrada, de fato, no Capítulo 8, com referência à discussão do teorema da função implícita.

EXERCÍCIO 7.6

1. Use determinantes jacobianos para testar a existência de dependência funcional entre os pares de funções:
 - (a) $y_1 = 3x_1^2 + x_2$
 $y_2 = 9x_1^4 + 6x_1^2(x_2 + 4) + x_2(x_2 + 8) + 12$
 - (b) $y_1 = 3x_1^2 + 2x_2^2$
 $y_2 = 5x_1 + 1$
2. Considere (7.22) como um conjunto de três funções $x_i^* = f^i(d_1, d_2, d_3)$ (com $i = 1, 2, 3$).
 - (a) Escreva o jacobiano 3×3 . Ele tem alguma relação com (7.23')? Podemos escrever $|J| = |V|$?
 - (b) Uma vez que $V \equiv (I - A)^{-1}$, podemos concluir que $|V| \neq 0$? O que podemos inferir disso em relação às três equações em (7.22)?

CAPÍTULO 8

Análise estática comparativa de modelos de função geral

O estudo de derivadas parciais no Capítulo 7 nos habilitou a manipular os tipos mais simples de problemas de estática comparativa, nos quais a solução de equilíbrio do modelo pode ser enunciada explicitamente na forma reduzida. Nesse caso, a diferenciação parcial da solução resultará diretamente na informação de estática comparativa desejada. Lembre-se de que a definição de derivada parcial requer a ausência de qualquer relação funcional entre as variáveis independentes (digamos, x_i), de modo que x_1 pode variar sem afetar os valores de $x_2, x_3, \dots, x_n$. Aplicada à análise estática comparativa, isso significa que os parâmetros e/ou variáveis exógenas que aparecem na forma reduzida da solução devem ser mutuamente independentes. Visto que, na verdade, eles são definidos como dados predeterminados para a finalidade do modelo, a possibilidade de eles afetarem mutuamente um ao outro está inerentemente excluída. O procedimento de diferenciação parcial adotado no Capítulo 7 é, portanto, totalmente justificável.

Contudo, não se deve esperar tal conveniência quando, devido à inclusão de funções gerais em um modelo, não seja possível obter nenhuma solução explícita em forma reduzida. Nesses casos, teremos de achar as derivadas estáticas comparativas diretamente das equações dadas originalmente no modelo. Tome, por exemplo, um modelo simples de renda nacional com duas variáveis endógenas Y e C :

$$Y = C + I_0 + G_0$$

$$C = C(Y, T_0) \quad [T_0: \text{impostos exógenos}]$$

que é redutível a uma única equação (uma condição de equilíbrio)

$$Y = C(Y, T_0) + I_0 + G_0$$

para ser resolvida para Y^* . Por causa da forma geral da função C , entretanto, não há nenhuma solução explícita disponível. Portanto, temos de achar as derivadas estáticas comparativas diretamente dessa equação. E como abordariamos esse problema? Qual dificuldade especial poderíamos encontrar?

Vamos supor que exista uma solução de equilíbrio Y^* . Então, sob certas condições muito gerais (que serão discutidas na Seção 8.5), podemos considerar Y^* como uma função diferenciável das variáveis exógenas I_0, G_0 e T_0 . Assim, podemos escrever a equação

$$Y^* = Y^*(I_0, G_0, T_0)$$

mesmo que não possamos determinar explicitamente a forma que a função assume. Além do mais, em alguma vizinhança do valor de equilíbrio Y^* , valerá a seguinte igualdade idêntica:

$$Y^* \equiv C(Y^*, T_0) + I_0 + G_0$$

Esse tipo de identidade será denominado uma *identidade de equilíbrio* porque nada mais é do que a condição de equilíbrio quando se substitui a variável Y por seu valor de equilíbrio Y^* . Agora que Y^* entrou em cena, poderá parecer, à primeira vista, que a simples diferenciação parcial dessa identidade resultará em alguma derivada estática comparativa, digamos, $\partial Y^* / \partial T_0$. Infelizmente, não é esse o caso. Visto que Y^* é uma função de T_0 , os dois argumentos da função C não são independentes. Especificamente, nesse caso, T_0 pode afetar C não apenas *diretamente*, mas também *indiretamente* via Y^* . Por conseqüência, a diferenciação parcial já não é mais adequada aos nossos propósitos. Então, como enfrentamos essa situação?

A resposta é que devemos recorrer à *diferenciação total* (em vez de à diferenciação parcial). Com base na noção de *diferenciais totais*, o processo de diferenciação total pode nos levar ao conceito relacionado de *derivada total*, que mede a taxa de variação de uma função tal como $C(Y^*, T_0)$ em relação ao argumento T_0 , quando T_0 também afeta o outro argumento, Y^* . Assim, uma vez familiarizados com esses conceitos, poderemos lidar com funções cujos argumentos não são todos independentes, e isso removeria o maior obstáculo que encontramos, até aqui, em nosso estudo da estática comparativa de um modelo de função geral. Como prelúdio à discussão desses conceitos, entretanto, temos de apresentar, antes de mais nada, a noção de *diferenciais*.

8.1 Diferenciais

Até aqui, o símbolo dy/dx para a derivada da função $y = f(x)$ tem sido considerado uma entidade isolada. Agora nós o reinterpretaremos como uma razão entre duas quantidades: dy e dx .

Diferenciais e derivadas

Por definição, a derivada $dy/dx = f'(x)$ é o limite de um quociente de diferenças:

$$\frac{dy}{dx} = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \quad (8.1)$$

Assim, por si só, $\Delta y / \Delta x$ (sem a exigência de $\Delta x \rightarrow 0$) não é igual a dy/dx . Se denotarmos a diferença entre os dois quocientes por δ , podemos escrever

$$\frac{\Delta y}{\Delta x} - \frac{dy}{dx} = \delta \quad \text{onde} \quad \delta \rightarrow 0 \quad \text{quando} \quad \Delta x \rightarrow 0 \quad [\text{por (8.1)}] \quad (8.2)$$

Multiplicando (8.2) por Δx e rearranjando os termos, temos:

$$\Delta y = \frac{dy}{dx} \Delta x + \delta \Delta x \quad \text{ou} \quad \Delta y = f'(x) \Delta x + \delta \Delta x \quad (8.3)$$

Essa equação descreve a variação em y (Δy) resultante de uma variação específica – não necessariamente pequena – em x (Δx) a partir de qualquer valor inicial de x no domínio da função $y = f(x)$. Mas ela também sugere que, ignorando a diferença $\delta \Delta x$, podemos utilizar o termo $f'(x) \Delta x$ como uma aproximação do valor verdadeiro de Δy , onde a aproximação fica progressivamente melhor à medida que Δx fica progressivamente menor.

Na Figura 8.1a, quando x varia de x_0 para $x_0 + \Delta x$, ocorre um movimento do ponto A até o ponto B no gráfico de $y = f(x)$. O Δy verdadeiro é medido pela distância CB , e a razão entre as duas distâncias $CB/AC = \Delta y / \Delta x$ pode ser determinada pela inclinação do segmento de reta AB . Mas, se desenharmos uma reta tangente AD no ponto A , e usarmos AD em vez de AB para aproxi-

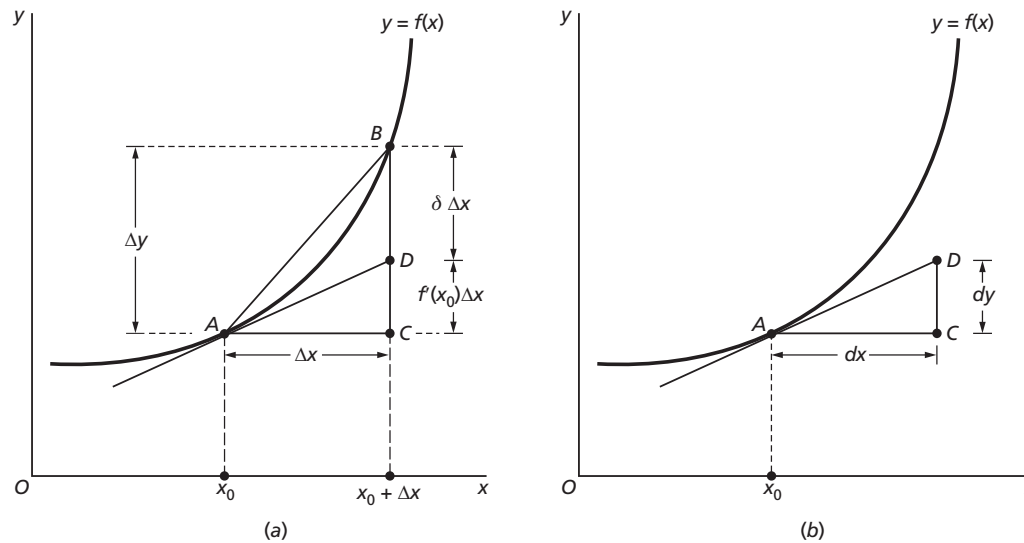


FIGURA 8.1

mar o valor de Δy , obtemos a distância CD ; a distância que sobra, DB , é a diferença ou o erro de aproximação. Visto que a inclinação de AD é $f'(x_0)$, a distância CD é igual a $f'(x_0) \Delta x$ e, por (8.3), a distância DB é igual a $\delta \Delta x$. Obviamente, à medida que Δx diminuir, o ponto B deslizará ao longo da curva em direção ao ponto A , reduzindo assim a diferença e tornando $f'(x)$ ou dy/dx uma aproximação melhor de $\Delta y/\Delta x$.

Focalizando a reta tangente AD , e tomando a distância CD como uma aproximação de CB , vamos renomear as distâncias AC e CD por dx e dy , respectivamente, como mostra a Figura 8.1b. Então,

$$\frac{dy}{dx} = \text{inclinação da tangente } AD = f'(x)$$

e, após multiplicar tudo por dx , obtemos

$$dy = f'(x) dx \quad (8.4)$$

Então, a derivada $f'(x)$ pode ser reinterpretada como o fator de proporcionalidade entre as duas variações finitas dy e dx . Conseqüentemente, dado um valor específico de dx , podemos multiplicá-lo por $f'(x)$ para obter dy como uma aproximação de Δy , entendendo que, quanto menor Δx , melhor a aproximação. As quantidades dx e dy são denominadas as *diferenciais* de x e y , respectivamente.

É preciso fazer algumas observações em relação a diferenciais como entidades matemáticas. Em primeiro lugar, enquanto dx é uma variável independente, dy é uma variável dependente. Especificamente, dy é uma função de x , bem como de dx : ela depende de x porque uma posição diferente de x_0 na Figura 8.1 significaria uma localização diferente para o ponto A e para sua reta tangente; ela depende de dx porque uma grandeza diferente de dx significaria uma posição diferente para o ponto C , bem como uma distância CD diferente. Em segundo lugar, se $dx = 0$, então $dy = 0$, porque, nesse caso, o ponto B coincidiria com o ponto A . Mas, se $dx \neq 0$, então é possível dividir dy por dx para obter $f'(x)$, exatamente como podemos multiplicar dx por $f'(x)$ para obter dy . Em terceiro lugar, a diferencial dy pode ser expressa somente em termos de alguma outra diferencial, ou diferenciais – aqui, dx . Isso porque nosso contexto exige a vinculação de uma variação dependente dy com uma variação independente dx . Embora tenha sentido escrever $dy = f'(x) dx$, não tem significado cortar o termo dx do lado direito e escrever $dy = f'(x)$. A vinculação das duas variações é efetuada por meio da derivada $f'(x)$, que pode ser vista como uma “conversora” que serve para traduzir uma dada variação dx em uma variação correspondente dy .

O processo de achar a diferencial dy a partir de uma dada função $y = f(x)$ é denominado *diferenciação*. Lembre-se de que temos utilizado esse termo como um sinônimo de derivação sem dar

uma explanação adequada. Em vista da nossa interpretação de uma derivada como um quociente de duas diferenciais, todavia, o princípio racional do termo fica evidente por si só. Porém, não deixa de ser ambíguo utilizar um único termo “diferenciação” para nos referirmos ao processo de achar a diferencial dy , bem como ao processo de achar a derivada dy/dx . Para evitar confusão, na prática, o usual é acrescentar à palavra *diferenciação* a expressão qualificativa “em relação a x ” quando tomamos a derivada dy/dx .

Diferenciais e elasticidade-ponto

Para ilustrar a aplicação de diferenciais à economia, vamos considerar a noção de elasticidade de uma função. Dada uma função demanda $Q = f(P)$, por exemplo, sua elasticidade é definida como $(\Delta Q/Q)/(\Delta P/P)$. Usando a idéia da aproximação explicada na Figura 8.1, podemos substituir a variação independente ΔP e a variação dependente ΔQ pelas diferenciais dP e dQ , respectivamente, para obter uma aproximação da medida da elasticidade conhecida como a *elasticidade-ponto* de demanda e denotada por ε_d (a letra grega épsilon, para “elasticidade”):[†]

$$\varepsilon_d \equiv \frac{dQ/Q}{dP/P} = \frac{dQ/dP}{Q/P} \quad (8.5)$$

Observe que, na extrema direita da expressão, as diferenciais dQ e dP foram rearranjadas como uma razão dQ/dP , que pode ser inferida como a derivada, ou a função *marginal*, da função demanda $Q = f(P)$. Uma vez que podemos interpretar, de maneira semelhante, a razão Q/P no denominador como a função *média* da função demanda, a elasticidade-ponto de demanda ε_d em (8.5) é considerada a razão entre a função marginal e a função média da função demanda.

De fato, essa relação que descrevemos por último é válida não somente para a função demanda mas também para qualquer outra função porque, para qualquer função *total* dada $y = f(x)$, podemos escrever a fórmula para a elasticidade-ponto de y em relação a x como:

$$\varepsilon_{yx} = \frac{dy/dx}{y/x} = \frac{\text{função marginal}}{\text{função média}} \quad (8.6)$$

Por questão de convenção, o valor *absoluto* da medida da elasticidade é usado ao decidir se a função é elástica em um determinado ponto. No caso de uma função demanda, por exemplo, estipulamos:

$$\text{A demanda é } \left\{ \begin{array}{l} \text{elástica} \\ \text{de elasticidade unitária} \\ \text{inelástica} \end{array} \right\} \text{ em um ponto quando } |\varepsilon_d| \gtrless 1.$$

EXEMPLO 1

Calcule ε_d se a função demanda for $Q = 100 - 2P$. A função marginal e a função média da demanda dada são

$$\frac{dQ}{dP} = -2 \quad \text{e} \quad \frac{Q}{P} = \frac{100 - 2P}{P}$$

portanto, a razão entre elas nos dá

$$\varepsilon_d = \frac{-P}{50 - P}$$

Essa fórmula mostra que a elasticidade é uma função de P . Contudo, tão logo seja escolhido um preço específico, a elasticidade-ponto terá uma grandeza determinada. Quando $P = 25$, por exemplo, temos $\varepsilon_d = -1$, ou $|\varepsilon_d| = 1$, de modo que a elasticidade de demanda é unitária naquele ponto. Já quando $P = 30$, temos $|\varepsilon_d| = 1,5$; por conseguinte, a demanda é elástica naquele preço.

[†] A medição da elasticidade-ponto pode ser interpretada alternativamente como o limite de $\frac{\Delta Q/Q}{\Delta P/P} = \frac{\Delta Q/\Delta P}{Q/P}$ quando $\Delta P \rightarrow 0$, o que dá o mesmo resultado que (8.5).

Em termos mais gerais, pode-se verificar que temos $|\varepsilon_d| > 1$ para $25 < P < 50$ e $|\varepsilon_d| < 1$ para $0 < P < 25$ no presente exemplo. (Um preço $P > 50$ pode ser considerado significativo, aqui?)

EXEMPLO 2

Ache a elasticidade-ponto de oferta ε_s a partir da função oferta $Q = P^2 + 7P$, e determine se a oferta é elástica em $P = 2$. Visto que as funções marginal e média são, respectivamente,

$$\frac{dQ}{dP} = 2P + 7 \quad \text{e} \quad \frac{Q}{P} = P + 7$$

a razão entre elas nos dá a elasticidade de oferta

$$\varepsilon_s = \frac{2P + 7}{P + 7}$$

Quadro $P = 2$, essa elasticidade é $11/9 > 1$; assim, a oferta é elástica em $P = 2$.

Ousando uma ligeira digressão, também podemos acrescentar aqui que a interpretação da razão entre duas diferenciais como uma derivada – e a conseqüente transformação da fórmula da elasticidade de uma função em uma razão entre sua função marginal e sua função média – possibilita um meio rápido de determinar a elasticidade-ponto graficamente. Os dois diagramas na Figura 8.2 ilustram, respectivamente, os casos de uma curva com inclinação negativa e de uma curva com inclinação positiva. Em cada caso, o valor da função marginal no ponto A da curva, ou em $x = x_0$ no domínio, é medido pela inclinação da reta tangente AB . O valor da função média, por outro lado, é medido, em cada caso, pela inclinação da reta OA (a reta que liga o ponto de origem com o ponto dado A na curva, como um vetor raio) porque no ponto A temos $y = x_0 A$ e $x = Ox_0$, de modo que a média é $y/x = x_0 A / Ox_0 =$ inclinação de OA . Assim, a elasticidade no ponto A pode ser prontamente verificada comparando os valores *numéricos* das duas inclinações envolvidas: se AB for mais íngreme que OA , a função é elástica no ponto A ; caso contrário, ela é inelástica em A . Por conseqüência, a função representada na Figura 8.2a é inelástica em A (ou em $x = x_0$), ao passo que a função na Figura 8.2b é elástica em A .

Além do mais, as duas inclinações comparadas são diretamente dependentes das dimensões respectivas dos dois ângulos θ_m e θ_a (letra grega teta; os índices indicam marginal e média, respectivamente). Assim, alternativamente, podemos comparar esses dois ângulos, em vez de comparar as duas inclinações correspondentes. Referindo-nos mais uma vez à Figura 8.2, podemos ver que $\theta_m < \theta_a$ no ponto A no diagrama *a*, o que indica que a marginal é menor que a média em valor numérico; assim, a função é inelástica no ponto A . O exato oposto é válido na Figura 8.2b.

Às vezes, estamos interessados em localizar um ponto de elasticidade unitária em uma dada curva. Isso pode ser feito com facilidade. Se a inclinação da curva for negativa, como na Figura 8.3a, devemos achar um ponto C tal que a reta OC e a tangente BC formem um ângulo do mesmo tamanho com o eixo x , embora na direção posta. No caso de uma curva com inclinação positiva, como na Figura 8.3b, basta achar um ponto C tal que a reta tangente em C , quando adequadamente estendida, passe pelo ponto de origem.

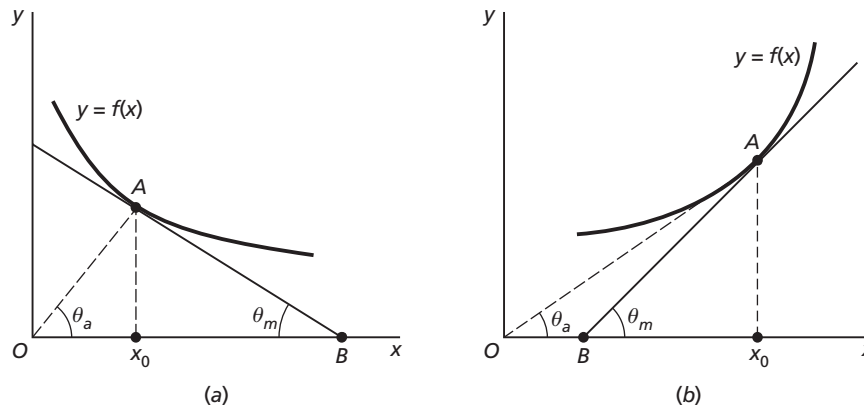
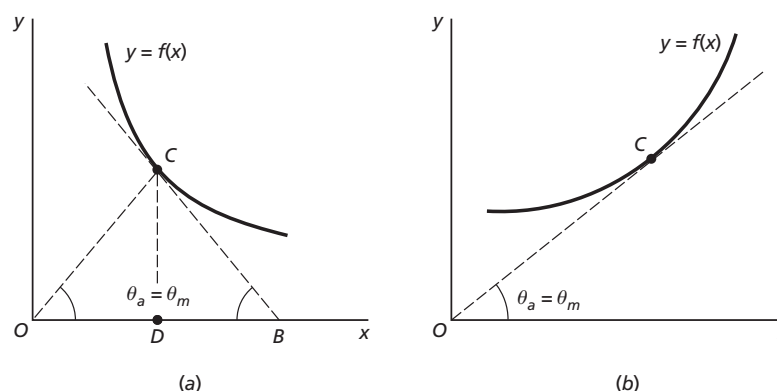


FIGURA 8.2

FIGURA 8.3



Devemos prevenir o leitor de que o método gráfico que acabamos de descrever baseia-se na premissa de que a função $y = f(x)$ é traçada com a variável dependente y no eixo vertical. Em particular, ao aplicar o método a uma curva de demanda, devemos nos certificar que Q esteja no eixo vertical. (Agora suponha que Q foi traçada no eixo horizontal. Como devemos modificar nosso método de leitura da elasticidade-ponto?)

EXERCÍCIO 8.1

- Calcule a diferencial dy , dadas:
 - $y = -x(x^2 + 3)$
 - $y = (x - 8)(7x + 5)$
 - $y = \frac{x}{x^2 + 1}$
- Dada a função importação $M = f(Y)$, onde M representa as importações e Y é a renda nacional, expresse a elasticidade-renda de importação ε_{MY} em termos das propensões a importar.
- Dada a função consumo $C = a + bY$ (com $a > 0$; $0 < b < 1$):
 - Calcule sua função marginal e sua função média.
 - Calcule a elasticidade-renda de consumo ε_{CY} e determine seu sinal, admitindo que $Y > 0$.
 - Mostre que essa função consumo é inelástica em todos os níveis positivos de renda.
- Calcule a elasticidade-ponto de demanda, dada $Q = k/P^n$, onde k e n são constantes positivas.
 - A elasticidade depende do preço, neste caso?
 - No caso especial em que $n = 1$, qual é o formato da curva de demanda? Qual é a elasticidade-ponto de demanda?
- Encontre uma curva de inclinação positiva cuja elasticidade-ponto seja constante em toda a extensão da curva.
 - Escreva a equação da curva e verifique, por (8.6), que a elasticidade é, de fato, uma constante.
- Dada $Q = 100 - 2P + 0,02Y$, onde Q é quantidade demandada, P é preço e Y é renda, e dados $P = 20$ e $Y = 5.000$, calcule:
 - A elasticidade-preço de demanda.
 - A elasticidade-renda de demanda.

8.2 Diferenciais totais

O conceito de diferenciais pode ser facilmente estendido a uma função de duas ou mais variáveis independentes. Considere uma função poupança

$$S = S(Y, i) \quad (8.7)$$

onde S é poupança, Y é renda nacional e i é taxa de juros. Admite-se que essa função – assim como todas as funções que usaremos aqui – é contínua e possui derivadas (parciais) contínuas ou, simbolicamente, $f \in C^1$. A derivada parcial $\partial S/\partial Y$ mede a propensão marginal a poupar. Assim, para qualquer variação em Y , dY , a variação resultante em S pode ser aproximada pela quantidade $(\partial S/\partial Y) dY$, que é comparável com a expressão do lado direito em (8.4). De modo semelhante, dada uma variação em i , di , podemos tomar $(\partial S/\partial i) di$ como a aproximação da variação resultante em S . A variação total em S é, então, aproximada pela diferencial

$$dS = \frac{\partial S}{\partial Y} dY + \frac{\partial S}{\partial i} di \quad (8.8)$$

ou, em uma notação alternativa,

$$dS = S_Y dY + S_i di$$

Note que as duas derivadas parciais S_Y e S_i novamente desempenham o papel de “conversoras” que servem para converter as variações dY e di , respectivamente, em uma variação correspondente dS . A expressão dS , por ser a soma das variações aproximadas de ambas as fontes, é denominada *diferencial total* da função poupança. E o processo de achar tal diferencial total é denominado *diferenciação total*. Ao contrário, os dois componentes aditivos do lado direito do sinal de igual em (8.8) são denominados *diferenciais parciais* da função poupança.

É claro que é possível Y variar enquanto i permanece constante. Nesse caso, $di = 0$, e a diferencial total se reduzirá a $dS = (\partial S/\partial Y) dY$. Dividindo ambos os lados por dY , obtemos:

$$\frac{\partial S}{\partial Y} = \left(\frac{dS}{dY} \right)_{i \text{ constante}}$$

Assim, fica claro que a derivada parcial $\partial S/\partial Y$ também pode ser interpretada, no espírito da Figura 8.1b, como a razão entre duas diferenciais dS e dY , com a condição de que i , a outra variável independente da função, seja mantida constante. Analogamente, podemos interpretar a derivada parcial $\partial S/\partial i$ como a razão entre a diferencial dS (com Y mantido constante) e a diferencial di . Note que, embora cada uma, dS e di , possa figurar sozinha como uma diferencial, a expressão $\partial S/\partial i$ permanece uma entidade única.

O caso mais geral de uma função de n variáveis independentes pode ser exemplificado por, digamos, uma função utilidade na forma geral

$$U = U(x_1, x_2, \dots, x_n) \quad (8.9)$$

A diferencial total dessa função pode ser representada por

$$dU = \frac{\partial U}{\partial x_1} dx_1 + \frac{\partial U}{\partial x_2} dx_2 + \dots + \frac{\partial U}{\partial x_n} dx_n \quad (8.10)$$

ou

$$dU = U_1 dx_1 + U_2 dx_2 + \dots + U_n dx_n = \sum_{i=1}^n U_i dx_i$$

na qual cada termo do lado direito indica a variação aproximada em U resultante de uma variação em uma das variáveis independentes. Em economia, o primeiro termo, $U_1 dx_1$, significa a utilidade marginal da primeira mercadoria vezes o incremento no consumo dessa mercadoria, sendo que o mesmo se aplica, de modo semelhante, aos outros termos. Assim, a soma desses termos, dU , representa a variação total aproximada na utilidade, originada de todas as fontes de variação possíveis. Como mostra o raciocínio em (8.3), dU , como uma aproximação, tende à verdadeira variação ΔU quando todos os termos dx_i tenderem a zero.

Como se pode esperar de qualquer outra função, ambas, a função poupança (8.7) e a função utilidade (8.9) dão origem a medidas de elasticidade-ponto semelhantes à definida em (8.6). Mas, nessas circunstâncias, cada medida de elasticidade deve ser definida em termos da variação em somente *uma* das variáveis independentes; assim, haverá *duas* dessas medidas de elasticidade para a função poupança, e *n* medidas para a função utilidade. Por isso, elas são denominadas *elasticidades parciais*. Para a função poupança, as elasticidades parciais podem ser escritas como

$$\varepsilon_{SY} = \frac{\partial S / \partial Y}{S/Y} = \frac{\partial S}{\partial Y} \frac{Y}{S} \quad \text{e} \quad \varepsilon_{Si} = \frac{\partial S / \partial i}{S/i} = \frac{\partial S}{\partial i} \frac{i}{S}$$

Para a função utilidade, as *n* elasticidades parciais podem ser denotadas, concisamente, como se segue:

$$\varepsilon_{Ux_i} = \frac{\partial U}{\partial x_i} \frac{x_i}{U} \quad (i = 1, 2, \dots, n)$$

EXEMPLO 1

Calcule a diferencial total para as seguintes funções utilidade, onde $a, b > 0$:

$$(a) U(x_1, x_2) = ax_1 + bx_2$$

$$(b) U(x_1, x_2) = x_1^2 + x_2^3 + x_1x_2$$

$$(c) U(x_1, x_2) = x_1^a x_2^b$$

As diferenciais totais são as seguintes:

$$(a) \quad \frac{\partial U}{\partial x_1} = U_1 = a \quad \frac{\partial U}{\partial x_2} = U_2 = b$$

e

$$dU = U_1 dx_1 + U_2 dx_2 = a dx_1 + b dx_2$$

$$(b) \quad \frac{\partial U}{\partial x_1} = U_1 = 2x_1 + x_2 \quad \frac{\partial U}{\partial x_2} = U_2 = 3x_2^2 + x_1$$

e

$$dU = U_1 dx_1 + U_2 dx_2 = (2x_1 + x_2) dx_1 + (3x_2^2 + x_1) dx_2$$

$$(c) \quad \frac{\partial U}{\partial x_1} = U_1 = ax_1^{a-1}x_2^b = \frac{ax_1^a x_2^b}{x_1} \quad \frac{\partial U}{\partial x_2} = U_2 = bx_1^a x_2^{b-1} = \frac{bx_1^a x_2^b}{x_2}$$

e

$$dU = \left(\frac{ax_1^a x_2^b}{x_1} \right) dx_1 + \left(\frac{bx_1^a x_2^b}{x_2} \right) dx_2$$

EXERCÍCIO 8.2

1. Expresse a diferencial total dU usando o vetor gradiente ∇U .

2. Calcule a diferencial total, dadas:

$$(a) z = 3x^2 + xy - 2y^3$$

$$(b) U = 2x_1 + 9x_1x_2 + x_2^2$$

3. Calcule a diferencial total, dadas:

$$(a) y = \frac{x_1}{x_1 + x_2} \quad (b) y = \frac{2x_1x_2}{x_1 + x_2}$$

4. A função oferta de uma certa mercadoria é
 $Q = a + bP^2 + R^{1/2}$ ($a < 0$, $b > 0$) [R : índice pluvial]
 Calcule a elasticidade-preço de oferta ε_{QP} e a elasticidade-índice pluvial de oferta ε_{QR} .
5. Como as duas elasticidades parciais no Problema 4 variam com P e R ? De um modo estritamente monotônico (admitindo P e R positivos)?
6. A demanda externa por nossas exportações, X , depende da renda externa Y_f e de nosso nível de preço P : $X = Y_f^{1/2} + P^{-2}$. Calcule a elasticidade parcial de demanda externa para nossas exportações em relação a nosso nível de preço.
7. Calcule a diferencial total para cada uma das seguintes funções:
 - (a) $U = -5x^3 - 12xy - 6y^5$
 - (b) $U = 7x^2 y^3$
 - (c) $U = 3x^3(8x - 7y)$
 - (d) $U = (5x^2 + 7y)(2x - 4y^3)$
 - (e) $U = \frac{9y^3}{x - y}$
 - (f) $U = (x - 3y)^3$

8.3 Regras de diferenciais

Um modo direto de achar a diferencial total dy , dada uma função

$$y = f(x_1, x_2)$$

é achar as derivadas parciais f_1 e f_2 e substituí-las na equação

$$dy = f_1 dx_1 + f_2 dx_2$$

Mas, às vezes, pode ser mais conveniente aplicar certas regras de diferenciais que, em vista de sua notável semelhança com as fórmulas de derivadas estudadas anteriormente, são muito fáceis de lembrar.

Seja k uma constante e u e v duas funções das variáveis x_1 e x_2 . Então, as seguintes regras são válidas:[†]

- | | | |
|------------------|---|--|
| Regra I | $dk = 0$ | (conforme regra da função constante) |
| Regra II | $d(cu^n) = cnu^{n-1} du$ | (conforme regra da função de potência) |
| Regra III | $d(u \pm v) = du \pm dv$ | (conforme regra da soma/diferença) |
| Regra IV | $d(uv) = v du + u dv$ | (conforme regra do produto) |
| Regra V | $d\left(\frac{u}{v}\right) = \frac{1}{v^2} (v du - u dv)$ | (conforme regra do quociente) |

Em vez de provar essas regras aqui, vamos apenas ilustrar sua aplicação prática.

[†] Todas as regras de diferenciais discutidas nesta seção também são aplicáveis quando as próprias u e v são as variáveis independentes (em vez de funções de algumas outras variáveis x_1 e x_2).

EXEMPLO 1 Calcule a diferencial total dy da função

$$y = 5x_1^2 + 3x_2$$

O método direto exige a avaliação das derivadas parciais $f_1 = 10x_1$ e $f_2 = 3$, que, então, nos habilitarão a escrever

$$dy = f_1 dx_1 + f_2 dx_2 = 10x_1 dx_1 + 3 dx_2$$

Contudo, podemos deixar que $u = 5x_1^2$ e $v = 3x_2$ e aplicar as regras dadas anteriormente para obter respostas idênticas, como segue:

$$dy = d(5x_1^2) + d(3x_2) \quad [\text{pela Regra III}]$$

$$= 10x_1 dx_1 + 3 dx_2 \quad [\text{pela Regra II}]$$

EXEMPLO 2 Calcule a diferencial total da função

$$y = 3x_1^2 + x_1x_2^2$$

Visto que $f_1 = 6x_1 + x_2^2$ e $f_2 = 2x_1x_2$, a diferencial desejada é

$$dy = (6x_1 + x_2^2) dx_1 + 2x_1x_2 dx_2$$

Aplicando as regras dadas, podemos chegar ao mesmo resultado assim:

$$dy = d(3x_1^2) + d(x_1x_2^2) \quad [\text{pela Regra III}]$$

$$= 6x_1 dx_1 + x_2^2 dx_1 + x_1 d(x_2^2) \quad [\text{pelas Regras II e IV}]$$

$$= (6x_1 + x_2^2) dx_1 + 2x_1x_2 dx_2 \quad [\text{pela Regra II}]$$

EXEMPLO 3 Calcule a diferencial total da função

$$y = \frac{x_1 + x_2}{2x_1^2}$$

Em vista do fato que, nesse caso, as derivadas parciais são

$$f_1 = \frac{-(x_1 + 2x_2)}{2x_1^3} \quad \text{e} \quad f_2 = \frac{1}{2x_1^2}$$

(verifique-as, como exercício), a diferencial desejada é

$$dy = \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2$$

Contudo, o mesmo resultado também pode ser obtido pela aplicação das regras da seguinte maneira:

$$dy = \frac{1}{4x_1^4} \left[2x_1^2 d(x_1 + x_2) - (x_1 + x_2) d(2x_1^2) \right] \quad [\text{pela Regra V}]$$

$$= \frac{1}{4x_1^4} [2x_1^2(dx_1 + dx_2) - (x_1 + x_2)4x_1dx_1] \quad [\text{pelas Regras III e II}]$$

$$= \frac{1}{4x_1^4} [-2x_1(x_1 + 2x_2)dx_1 + 2x_1^2 dx_2]$$

$$= \frac{-(x_1 + 2x_2)}{2x_1^3} dx_1 + \frac{1}{2x_1^2} dx_2$$

Essas regras podem ser naturalmente estendidas a casos em que estão envolvidas mais de duas funções de x_1 e x_2 . Em particular, podemos acrescentar as duas regras seguintes à coleção anterior:

Regra VI $d(u \pm v \pm w) = du \pm dv \pm dw$

Regra VII $d(uvw) = vw du + uw dv + uv dw$

Para derivar a Regra VII, podemos empregar o estratagema familiar de primeiramente assumir que $z = vw$, de modo que

$$d(uvw) = d(uz) = z du + u dz \quad [\text{pela Regra IV}]$$

Então, aplicando novamente a Regra IV a dz , obtemos o resultado intermediário

$$dz = d(vw) = w dv + v dw$$

que, quando substituído na equação precedente, resultará em

$$d(uvw) = vw du + u(w dv + v dw) = vw du + uw dv + uv dw$$

como o resultado final desejado. Um procedimento semelhante pode ser empregado para obter a Regra VI.

EXERCÍCIO 8.3

- Use as regras de diferenciais para calcular (a) dz de $z = 3x^2 + xy - 2y^3$ e (b) dU de $U = 2x_1 + 9x_1x_2 + x_2^2$. Compare suas respostas com as obtidas no Exercício 8.2-2.
- Use as regras de diferenciais para calcular dy das seguintes funções:
 (a) $y = \frac{x_1}{x_1 + x_2}$ (b) $y = \frac{2x_1x_2}{x_1 + x_2}$
 Compare suas respostas com as obtidas no Exercício 8.2-3.
- Dado $y = 3x_1(2x_2 - 1)(x_3 + 5)$
 (a) Calcule dy pela Regra VII.
 (b) Calcule a diferencial de y , se $dx_2 = dx_3 = 0$.
- Demonstre as Regras II, III, IV, e V, admitindo que u e v sejam as variáveis independentes (em vez de funções de algumas outras variáveis).

8.4 Derivadas totais

Agora atacaremos a questão proposta no início do capítulo; a saber, como podemos calcular a taxa de variação da função $C(Y^*, T_0)$ em relação a T_0 , quando Y^* e T_0 são relacionadas. Como já mencionamos anteriormente, a resposta está no conceito de derivada total. Diferente de uma derivada *parcial*, uma derivada *total* não exige que o argumento Y^* permaneça constante enquanto T_0 varia e, assim, pode levar em conta a relação entre os dois argumentos.

Calculando a derivada total

Para continuar a discussão segundo uma estrutura geral, vamos considerar qualquer função

$$y = f(x, w) \quad \text{onde} \quad x = g(w) \quad (8.11)$$

As duas funções f e g também podem ser combinadas em uma função composta

$$y = f[g(w), w] \quad (8.11')$$

As três variáveis y , x e w são relacionadas entre si, como mostra a Figura 8.4. Nessa figura, que denominaremos um *mapa de canal*, pode-se ver claramente que w – a fonte de variação definitiva – pode afetar y por dois canais distintos: (1) *indiretamente*, via a função g e então f (as setas retas); e (2) *diretamente*, via a função f (a seta curva). O efeito direto pode ser representado simplesmente pela derivada parcial f_w . Mas o efeito indireto somente pode ser expresso por um produto de duas derivadas, $f_x \frac{dx}{dw}$, ou $\frac{\partial y}{\partial x} \frac{dx}{dw}$, pela regra da cadeia para uma função composta. Adicionando os dois efeitos, temos a derivada total de y em relação a w que desejamos:

$$\begin{aligned} \frac{dy}{dw} &= f_x \frac{dx}{dw} = f_w \\ &= \frac{\partial y}{\partial x} \frac{dx}{dw} + \frac{\partial y}{\partial w} \end{aligned} \quad (8.12)$$

Essa derivada total também pode ser obtida por um método alternativo: podemos, em primeiro lugar, diferenciar totalmente a função $y = f(x, w)$ para obter a diferencial total

$$dy = f_x dx + f_w dw$$

e então dividir tudo por dw . O resultado é idêntico ao de (8.12). Seja qual for o método, denominamos o processo de achar a derivada total dy/dw *diferenciação total de y em relação a w* .

É de extrema importância distinguir entre os dois símbolos parecidos dy/dw e $\partial y/\partial w$ em (8.12). O primeiro é uma derivada *total* e o último, uma derivada *parcial*. Esta última é, na verdade, um mero componente da primeira.

EXEMPLO 1 Calcule a derivada total dy/dw , dada a função

$$y = f(x, w) = 3x - w^2 \quad \text{onde} \quad x = g(w) = 2w^2 + w + 4$$

Em virtude de (8.12), a derivada total deve ser

$$\frac{dy}{dw} = 3(4w + 1) + (-2w) = 10w + 3$$

À guisa de verificação, podemos substituir a função g na função f , para obter

$$y = 3(2w^2 + w + 4) - w^2 = 5w^2 + 3w + 12$$

que agora é uma função somente de w . Então, descobre-se facilmente que a derivada dy/dw é $10w + 3$, a resposta idêntica.

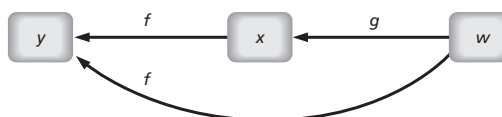
EXEMPLO 2 Se tivermos uma função utilidade $U = U(c, s)$, onde c é a quantidade de café consumida e s é a quantidade de açúcar consumido e uma outra função $s = g(c)$ que indica a complementaridade entre essas duas mercadorias, então podemos escrever simplesmente a função composta

$$U = U[c, g(c)]$$

da qual decorre que

$$\frac{dU}{dc} = \frac{\partial U}{\partial c} + \frac{\partial U}{\partial g(c)} g'(c)$$

FIGURA 8.4



Uma variação sobre o tema

A situação é apenas ligeiramente mais complicada quando temos

$$y = f(x_1, x_2, w) \quad \text{onde} \quad \begin{cases} x_1 = g(w) \\ x_2 = h(w) \end{cases} \quad (8.13)$$

Agora, o aspecto do mapa de canal será o apresentado na Figura 8.5. Desta vez, a variável w pode afetar y por três canais: (1) indiretamente, via a função g e então f ; (2) mais uma vez, indiretamente, via a função h e então f ; e (3) diretamente via f . Pela nossa experiência prévia, esperamos que esses efeitos possam ser expressos respectivamente como $\frac{\partial y}{\partial x_1} \frac{dx_1}{dw}$, $\frac{\partial y}{\partial x_2} \frac{dx_2}{dw}$ e $\frac{\partial y}{\partial w}$. Somando

tudo, obtemos a derivada total

$$\begin{aligned} \frac{dy}{dw} &= \frac{\partial y}{\partial x_1} \frac{dx_1}{dw} + \frac{\partial y}{\partial x_2} \frac{dx_2}{dw} + \frac{\partial y}{\partial w} \\ &= f_1 \frac{dx_1}{dw} + f_2 \frac{dx_2}{dw} + f_w \end{aligned} \quad (8.14)$$

que é análoga a (8.12). Se tomarmos a diferencial total dy , e, então, dividirmos tudo por dw , podemos chegar ao mesmo resultado.

EXEMPLO 3 Seja a função produção

$$Q = Q(K, L, t)$$

onde, à parte os dois insumos K e L , há um terceiro argumento t , que denota o tempo. A presença do argumento t indica que a função produção pode mudar ao longo do tempo, refletindo mudanças tecnológicas. Assim, essa é uma função produção dinâmica, e não estática. Uma vez que capital e mão-de-obra também podem variar ao longo do tempo, podemos escrever

$$K = K(t) \quad \text{e} \quad L = L(t)$$

Então, a taxa de variação do produto em relação ao tempo pode ser expressa, segundo a fórmula da derivada total (8.14), como

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial K} \frac{dK}{dt} + \frac{\partial Q}{\partial L} \frac{dL}{dt} + \frac{\partial Q}{\partial t}$$

ou, em uma notação alternativa,

$$\frac{dQ}{dt} = Q_K K'(t) + Q_L L'(t) + Q_t$$

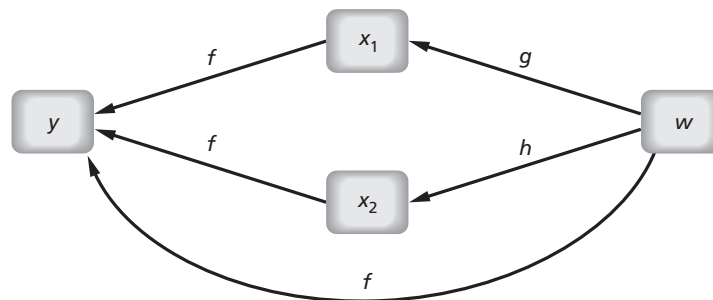


FIGURA 8.5

Mais uma variação sobre o tema

Quando a fonte definitiva de variação, w em (8.13), é substituída por duas fontes coexistentes, u e v , a situação se torna a seguinte:

$$y = f(x_1, x_2, u, v) \quad \text{onde} \quad \begin{cases} x_1 = g(u, v) \\ x_2 = h(u, v) \end{cases} \quad (8.15)$$

Embora agora o mapa de canal contenha mais setas, o princípio de sua construção continua o mesmo; portanto, deixaremos que você o desenhe. Para obter a derivada total de y em relação a u (enquanto v é mantida constante), vamos tomar a diferencial total de y , e então dividir tudo pela diferencial du , com o seguinte resultado:

$$\begin{aligned} \frac{dy}{du} &= \frac{\partial y}{\partial x_1} \frac{dx_1}{du} + \frac{\partial y}{\partial x_2} \frac{dx_2}{du} + \frac{\partial y}{\partial u} \frac{du}{du} + \frac{\partial y}{\partial v} \frac{dv}{du} \\ &= \frac{\partial y}{\partial x_1} \frac{dx_1}{du} + \frac{\partial y}{\partial x_2} \frac{dx_2}{du} + \frac{\partial y}{\partial u} \quad \left[\frac{dv}{du} = 0 \text{ desde que } v \text{ seja mantida constante} \right] \end{aligned}$$

Entretanto, em vista do fato de estarmos variando u enquanto mantemos v constante (pois uma única derivada não comporta variações em ambas, u e v), o resultado obtido deve ser modificado de duas maneiras: (1) as derivadas dx_1/du e dx_2/du do lado direito devem ser reescritas com o sinal de derivada parcial como $\partial x_1/\partial u$ e $\partial x_2/\partial u$, que está de acordo com as funções g e h em (8.15); e (2) a razão dy/du à esquerda também deve ser interpretada como uma derivada *parcial*, mesmo que – por ser derivada por meio do processo de diferenciação total de y – ela tenha, na verdade, as características de uma derivada *total*. Por essa razão, nos referiremos a ela pelo nome explícito de *derivada parcial total* e a denotaremos $\S y/\S u$ (usando $\S$ em vez de ∂), para distingui-la da derivada parcial simples $\partial y/\partial u$ que, como nosso resultado mostra, nada mais é do que um dos três termos componentes que, somados, resultam na derivada parcial total.[†]

Com essas modificações, nosso resultado torna-se

$$\frac{\S y}{\S u} = \frac{\partial y}{\partial x_1} \frac{\partial x_1}{\partial u} + \frac{\partial y}{\partial x_2} \frac{\partial x_2}{\partial u} + \frac{\partial y}{\partial u} \quad (8.16)$$

que é análogo a (8.14). Note que apareceu o símbolo $\partial y/\partial u$ do lado direito, o que torna necessária a adoção do novo símbolo $\S y/\S u$ do lado esquerdo para indicar o conceito mais amplo de uma derivada parcial total. De um modo perfeitamente análogo, podemos derivar a outra derivada parcial total, $\S y/\S v$. Contudo, considerando que os papéis de u e v são simétricos em (8.15), há uma outra alternativa mais simples, disponível para nós. Para obter $\S y/\S v$, basta substituir o símbolo u em (8.16) pelo símbolo v em tudo.

A utilização dos novos símbolos $\S y/\S u$ e $\S y/\S v$ para as derivadas parciais totais, mesmo que não convencional, cumpre a boa finalidade de evitar confusão com as derivadas parciais simples $\partial y/\partial u$ e $\partial y/\partial v$ que podem surgir somente da função f em (8.15). Contudo, no caso especial em que a função f toma a forma de $y = f(x_1, x_2)$, sem os argumentos u e v , as derivadas parciais simples $\partial y/\partial u$ e $\partial y/\partial v$ não são definidas. Por conseguinte, nesse caso, talvez não seja inadequado utilizar estes últimos símbolos para as derivadas parciais totais de y em relação a u e v , visto que provavelmente não surgirá nenhuma confusão. Ainda assim, é aconselhável utilizar um símbolo especial por questão de maior clareza.

[†] Um modo alternativo de denotar essa derivada parcial total é

$$\left. \frac{dy}{du} \right|_v \text{ constante} \quad \text{ou} \quad \left. \frac{dy}{du} \right|_{dv=0}$$

Algumas observações gerais

Para concluir esta seção, apresentamos três observações gerais referentes à derivada total e à diferenciação total.

1. Nos casos que discutimos, a situação envolve, sem exceção, uma variável que é funcionalmente dependente de uma segunda variável, que, por sua vez, depende funcionalmente de uma terceira variável. Como consequência, a noção de uma *cadeia* entra inevitavelmente em cena, como evidencia a aparição de um produto (ou produtos) de duas expressões derivadas como componente(s) de uma derivada total. Por essa razão, as fórmulas de derivadas totais em (8.12), (8.14) e (8.16) também podem ser consideradas como expressões da regra da cadeia, ou da regra da função composta – uma versão mais sofisticada da regra da cadeia apresentada na Seção 7.3.
2. A cadeia de derivadas não tem de ser limitada a apenas dois “elos” (a multiplicação de duas derivadas); o conceito de derivada total deve ser estendido aos casos em que há três ou mais elos na função composta.
3. Em todos os casos discutidos, as derivadas totais – incluindo as que foram denominadas *derivadas parciais totais* – medem taxas de variação em relação a algumas variáveis *definitivas* na cadeia ou, em outras palavras, em relação a certas variáveis que são, em certo sentido, *exógenas* e que *não* são expressas como funções de outras variáveis. A essência da derivada total e do processo de diferenciação total é dar a devida consideração a *todos* os canais, indiretos e diretos, por meio dos quais os efeitos de uma variação sobre uma variável independente *definitiva* possa ser transmitido à variável específica em estudo.

EXERCÍCIO 8.4

1. Calcule a derivada total dz/dy , dadas:
 - (a) $z = f(x, y) = 5x + xy - y^2$, onde $x = g(y) = 3y^2$
 - (b) $z = 4x^2 - 3xy + 2y^2$, onde $x = 1/y$
 - (c) $z = (x + y)(x - 2y)$, onde $x = 2 - 7y$
2. Calcule a derivada total dz/dt , dadas:
 - (a) $z = x^2 - 8xy - y^3$, onde $x = 3t$ e $y = 1 - t$
 - (b) $z = 7u + vt$, onde $u = 2t^2$ e $v = t + 1$
 - (c) $z = f(x, y, t)$, onde $x = a + bt$ e $y = c + kt$
3. Calcule a taxa de variação de produção em relação ao tempo, se a função produção for $Q = A(t)K^\alpha L^\beta$, onde $A(t)$ é uma função decrescente de t , e $K = K_0 + at$, e $L = L_0 + bt$.
4. Calcule as derivadas totais parciais $\$W/\u e $\$W/\v se:
 - (a) $W = ax^2 + bxy + cu$, onde $x = \alpha u + \beta v$ e $y = \gamma u$
 - (b) $W = f(x_1, x_2)$, onde $x_1 = 5u^2 + 3v$ e $x_2 = u - 4v^3$
5. Desenhe um mapa de canal adequado ao caso de (8.15).
6. Obtenha a expressão para $\$y/\v formalmente a partir de (8.15) tomando a diferencial total de y e então dividindo tudo por dv .

8.5 Derivadas de funções implícitas

O conceito de diferenciais totais também pode nos habilitar a calcular as derivadas das denominadas funções implícitas.

Funções implícitas

Uma função dada na forma $y = f(x)$, digamos,

$$y = f(x) = 3x^4 \quad (8.17)$$

é denominada uma *função explícita*, porque a variável y é explicitamente expressa como uma função de x . Se essa função for escrita alternativamente na forma equivalente

$$y - 3x^4 = 0 \quad (8.17')$$

contudo, não teremos mais uma função explícita. Ao contrário, (8.17) é, então, apenas *implicitamente* definida pela equação (8.17'). Portanto, quando temos (apenas) uma equação na forma de (8.17'), a função $y = f(x)$ que ela implica e cuja forma específica pode até mesmo não ser conhecida por nós, é denominada uma *função implícita*.

Uma equação na forma de (8.17') pode ser denotada, em geral, por $F(y, x) = 0$, porque seu lado esquerdo é uma função das duas variáveis y e x . Note que aqui estamos usando a letra maiúscula F para distingui-la da função f ; a função F , que representa o lado esquerdo da expressão em (8.17'), tem dois argumentos, y e x , ao passo que a função f , que representa a função implícita, tem somente um argumento, x . É claro que pode haver mais de dois argumentos na função F . Por exemplo, podemos encontrar uma equação $F(y, x_1, \dots, x_m) = 0$. Tal equação também *pode* definir uma função implícita $y = f(x_1, \dots, x_m)$.

Nesta última frase, a palavra ambígua “*pode*” foi usada propositalmente. Pois, enquanto uma função explícita, digamos, $y = f(x)$, sempre pode ser transformada em uma equação $F(y, x) = 0$ simplesmente transferindo a expressão $f(x)$ para o lado esquerdo do sinal de igualdade, a transformação inversa nem sempre é possível. De fato, em certos casos, uma dada equação, na forma de $F(y, x) = 0$ pode não definir implicitamente uma função $y = f(x)$. Por exemplo, a equação $x^2 + y^2 = 0$ é satisfeita somente no ponto de origem $(0, 0)$ e, por conseguinte, não resulta em nenhuma função que valha a pena mencionar. Como outro exemplo, a equação

$$F(y, x) = x^2 + y^2 - 9 = 0 \quad (8.18)$$

não implica uma função, mas uma relação, porque a representação gráfica de (8.18) é um círculo, como ilustra a Figura 8.6, de modo que nenhum valor único de y corresponde a cada valor de x . Entretanto, note que, se restringirmos y a valores não-negativos, então teremos somente a metade superior do círculo e isso constitui uma função, a saber, $y = +\sqrt{9 - x^2}$. De modo semelhante, a metade inferior do círculo com valores de y não-positivos constitui uma outra função, $y = -\sqrt{9 - x^2}$. Em contraste, nem a metade esquerda nem a metade direita do círculo podem se qualificar como uma função.

Em vista dessa incerteza, torna-se interessante indagar se há condições gerais conhecidas segundo as quais podemos ter certeza que uma dada equação na forma de

$$F(y, x_1, \dots, x_m) = 0 \quad (8.19)$$

realmente define uma função implícita

$$y = f(x_1, \dots, x_m) \quad (8.20)$$

localmente, isto é, ao redor de algum ponto específico no domínio. A resposta a essa pergunta está no denominado teorema da função implícita, cujo enunciado é:

Dada (8.19), se (a) a função F tiver derivadas parciais contínuas $F_y, F_{x_1}, \dots, F_{x_m}$, e se (b) em um ponto $(y_0, x_{10}, \dots, x_{m0})$ que satisfaça a equação (8.19), F_y for diferente de zero, então existe uma vizinhança m -dimensional de $(x_{10}, \dots, x_{m0})$, N , na qual y é uma função implicitamente definida das variáveis $x_1, \dots, x_m$, na forma de (8.20). Essa função implícita satisfaz $y_0 = f(x_{10}, \dots, x_{m0})$. Ela satisfaz também a equação (8.19) para *toda* m -upla $(x_1, \dots, x_m)$ na vizinhança de N —dando a (8.19), por conseguinte, o status de uma *identidade* naquela vizinhança. Além do mais, a função implícita f é contínua e tem derivadas parciais contínuas $f_1, \dots, f_m$.

Vamos aplicar esse teorema à equação do círculo, (8.18), que contém somente uma variável x . Em primeiro lugar, podemos verificar devidamente se $F_y = 2y$ e $F_x = 2x$ são contínuas, como requerido. Então, notamos que F_y é diferente de zero exceto quando $y = 0$, isto é, exceto no ponto mais à esquerda $(-3, 0)$ e no ponto mais à direita $(3, 0)$ no círculo. Assim, ao redor de qualquer ponto no círculo, exceto $(-3, 0)$ e $(3, 0)$, podemos construir uma vizinhança na qual a equação (8.18) define uma função implícita $y = f(x)$. Isso pode ser verificado com facilidade na Figura 8.6, onde, de fato, é possível desenhar, digamos, um retângulo ao redor de qualquer ponto do círculo – exceto em $(-3, 0)$ e $(3, 0)$ – de modo que a porção do círculo contida no retângulo constituirá o gráfico de uma função com um valor único de y exclusivo para cada valor de x naquele retângulo.

Há diversas coisas que devem ser notadas no teorema da função implícita. Em primeiro lugar, as condições citadas no teorema têm as características de condições suficientes (mas não necessárias). Isso significa que, se acaso acharmos $F_y = 0$ em um ponto que satisfaça (8.19), não podemos usar o teorema para negar a existência de uma função implícita ao redor desse ponto. Pois tal função pode, de fato, existir (ver Exercícios 8.5 a 8.7).[†] Em segundo lugar, mesmo que tenhamos certeza de que existe uma função f , o teorema não dá nenhuma indicação quanto à forma específica que a função f assume e, a propósito, tampouco nos informa o tamanho exato da vizinhança N na qual a função implícita é definida. Contudo, apesar dessas limitações, esse teorema é de grande importância, pois, sempre que as condições nele enunciadas forem satisfeitas, torna-se significativo falar sobre uma função e utilizá-la como a (8.20), mesmo que nosso modelo possa conter uma equação (8.19), que é difícil ou impossível de resolver explicitamente para y em termos das variáveis x . Além disso, uma vez que o teorema também garante a existência das derivadas parciais $f_1, \dots, f_m$, agora também tem significado falar sobre essas derivadas da função implícita.

Derivadas de funções implícitas

Se a equação $F(y, x_1, \dots, x_m) = 0$ puder ser resolvida para y , podemos escrever explicitamente a função $y = f(x_1, \dots, x_m)$ e obter suas derivadas pelos métodos que aprendemos antes. Por exemplo, (8.18) pode ser resolvida para resultar em duas funções separadas:

$$\begin{aligned} y^+ &= +\sqrt{9-x^2} && \text{[metade superior do círculo]} \\ y^- &= -\sqrt{9-x^2} && \text{[metade inferior do círculo]} \end{aligned} \quad (8.18')$$

e suas derivadas podem ser obtidas como se segue:

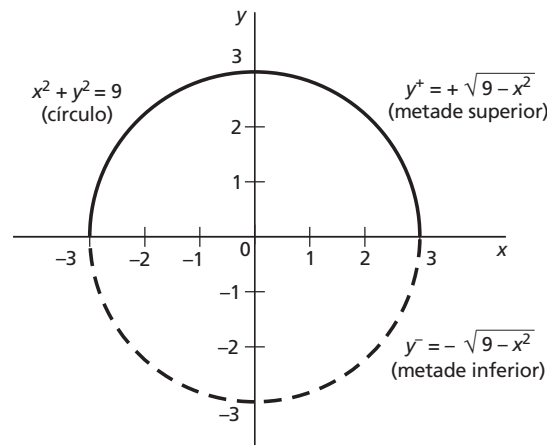


FIGURA 8.6

[†] Por outro lado, se $F_y = 0$ for uma vizinhança inteira, então pode-se concluir que nenhuma função implícita é definida naquela vizinhança. Pelo mesmo critério, se $F_y = 0$ identicamente, então não existe nenhuma função implícita em nenhum lugar.

$$\begin{aligned}
 \frac{dy^+}{dx} &= \frac{d}{dx} (9 - x^2)^{1/2} = \frac{1}{2} (9 - x^2)^{-1/2} (-2x) \\
 &= \frac{-x}{\sqrt{9 - x^2}} = \frac{-x}{y^+} \quad (y^+ \neq 0) \\
 \frac{dy^-}{dx} &= \frac{d}{dx} [-(9 - x^2)^{1/2}] = -\frac{1}{2} (9 - x^2)^{-1/2} (-2x) \\
 &= \frac{x}{\sqrt{9 - x^2}} = \frac{-x}{y^-} \quad (y^- \neq 0)
 \end{aligned} \tag{8.21}$$

Mas, e se a equação dada, $F(y, x_1, \dots, x_m) = 0$, não puder ser resolvida explicitamente para y ? Nesse caso, se soubermos que existe uma função implícita conforme os termos do teorema da função implícita, ainda podemos obter as derivadas desejadas sem ter de resolver para y primeiro. Para fazer isso, utilizamos a denominada regra da função implícita – uma regra que pode nos dar as derivadas de *todas* as funções implícitas definidas pela equação dada. O desenvolvimento dessa regra depende dos seguintes fatos básicos: (1) se duas expressões forem *identicamente* iguais, suas respectivas diferenciais totais devem ser iguais;[†] (2) a diferenciação de uma expressão que envolve $y, x_1, \dots, x_m$ resultará em uma expressão que envolve as diferenciais $dy, dx_1, \dots, dx_m$; e (3) a diferencial de y, dy , pode ser substituída, de modo que o fato de não podermos resolvê-la para y não tem importância.

Aplicando esses fatos à equação $F(y, x_1, \dots, x_m) = 0$ – a qual, vale lembrar, tem o status de uma *identidade* na vizinhança N na qual a função implícita é definida – podemos escrever $dF = d0$, ou

$$F_y dy + F_1 dx_1 + F_2 dx_2 + \dots + F_m dx_m = 0 \tag{8.22}$$

Visto que a função implícita $y = f(x_1, x_2, \dots, x_m)$ tem a diferencial total

$$dy = f_1 dx_1 + f_2 dx_2 + \dots + f_n dx_n$$

podemos substituir essa expressão de dy em (8.22) para obter (após reunir termos)

$$(F_y f_1 + F_1) dx_1 + (F_y f_2 + F_2) dx_2 + \dots + (F_y f_m + F_m) dx_m = 0 \tag{8.22'}$$

O fato de que todas as dx_i podem variar independentemente umas das outras significa que, para que a equação (8.22') seja válida, cada expressão entre parênteses tem de desaparecer individualmente, isto é, devemos ter

[†] Tome, por exemplo, a identidade

$$x^2 - y^2 \equiv (x + y)(x - y)$$

Isso é uma identidade porque os dois lados são iguais para *quaisquer* valores que queiramos atribuir a x e y . Tomando a diferencial total de cada lado, temos

$$\begin{aligned}
 d(\text{lado esquerdo}) &= 2x dx - 2y dy \\
 d(\text{lado direito}) &= (x - y) d(x + y) + (x + y) d(x - y) \\
 &= (x - y)(dx + dy) + (x + y)(dx - dy) \\
 &= 2x dx - 2y dy
 \end{aligned}$$

Os dois resultados são, de fato, iguais. Contudo, se as duas expressões *não* forem identicamente iguais, mas iguais apenas para certos valores específicos das variáveis, suas diferenciais totais *não* serão iguais. A equação

$$x^2 - y^2 = x^2 + y^2 - 2$$

por exemplo, é válida somente para $y = \pm 1$. As diferenciais totais dos dois lados são

$$\begin{aligned}
 d(\text{lado esquerdo}) &= 2x dx - 2y dy \\
 d(\text{lado direito}) &= 2x dx + 2y dy
 \end{aligned}$$

que não são iguais. Note, em particular, que elas não são iguais nem mesmo para $y = \pm 1$.

$$F_y f_i + F_i = 0 \quad (\text{para todo } i)$$

Dividindo tudo por F_y e resolvendo para f_i , obtemos a denominada regra da função implícita para achar a derivada parcial f_i da função implícita $y = f(x_1, x_2, \dots, x_m)$:

$$f_i \equiv \frac{\partial y}{\partial x_i} = -\frac{F_i}{F_y} \quad (i = 1, 2, \dots, m) \quad (8.23)$$

No caso simples em que a equação dada é $F(y, x) = 0$, a regra nos dá

$$\frac{dy}{dx} = -\frac{F_x}{F_y} \quad (8.23')$$

O que essa regra mostra é que, mesmo que a forma específica da função implícita não seja conhecida, podemos, não obstante, achar sua(s) derivada(s) tomando a *negativa* da razão de um par de derivadas parciais da função F que aparece na equação dada que define a função implícita. Observe que F_y sempre aparece no denominador da razão. Sendo esse o caso, não é admissível ter $F_y = 0$. Posto que o teorema da função implícita especifica que $F_y \neq 0$ no ponto ao redor do qual a função implícita é definida, o problema de um denominador zero está automaticamente resolvido na vizinhança relevante daquele ponto.

EXEMPLO 1

Calcule dy/dx para a função implícita definida por (8.17'). Visto que $F(y, x)$ toma a forma de $y - 3x^4$, temos, por (8.23'),

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{-12x^3}{1} = 12x^3$$

Neste caso particular, podemos resolver facilmente a equação dada para y para obter $y = 3x^4$. Assim, a correção da derivada é verificada com facilidade.

EXEMPLO 2

Calcule dy/dx para as funções implícitas definidas pela equação do círculo (8.18). Desta vez, temos $F(y, x) = x^2 + y^2 - 9$; assim, $F_y = 2y$ e $F_x = 2x$. Por (8.23'), a derivada desejada é

$$\frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y} \quad (y \neq 0)^{\dagger}$$

Afirmamos, anteriormente, que a regra da função implícita nos dá a derivada de *toda* função implícita definida por uma dada equação. Vamos verificar essa afirmativa com as duas funções em (8.18') e suas derivadas em (8.21). Se substituirmos y^+ por y no resultado da regra da função implícita $dy/dx = -x/y$, de fato obteremos a derivada dy^+/dx , como mostrado em (8.21); de modo semelhante, a substituição de y^- por y resultará na outra derivada em (8.21). Assim, nossa afirmativa anterior está devidamente verificada.

EXEMPLO 3

Calcule $\partial y/\partial x$ para qualquer função (ou quaisquer funções) implícita(s) que possa(m) ser definida(s) pela equação $F(y, x, w) = y^3 x^2 + w^3 + yxw - 3 = 0$. Essa equação não é resolvida facilmente para y . Mas, visto que F_y , F_x , e F_w são todas obviamente contínuas e, posto que $F_y = 3y^2 x^2 + xw$ é, de fato, diferente de zero em um ponto tal como $(1, 1, 1)$, que satisfaz a equação dada, certamente existe uma função implícita $y = f(x, w)$ ao menos ao redor daquele ponto. Portanto, tem sentido falar sobre a derivada $\partial y/\partial x$. Além disso, por (8.23), podemos escrever imediatamente:

$$\frac{\partial y}{\partial x} = -\frac{F_x}{F_y} = -\frac{2y^3 x + yw}{3y^2 x^2 + xw}$$

No ponto $(1, 1, 1)$, o valor dessa derivada é $-\frac{3}{4}$.

[†] É evidente que a restrição $y \neq 0$ é perfeitamente coerente com nossa discussão anterior da equação (8.18) que segue o enunciado do teorema da função implícita.

EXEMPLO 4

Admita que a equação $F(Q, K, L) = 0$ define implicitamente uma função de produção $Q = f(K, L)$. Vamos achar um meio de expressar os produtos marginais físicos MPP_K e MPP_L em relação à função F . Visto que os produtos marginais são simplesmente as derivadas parciais $\partial Q / \partial K$ e $\partial Q / \partial L$, podemos aplicar a regra da função implícita e escrever

$$MPP_K \equiv \frac{\partial Q}{\partial K} = -\frac{F_K}{F_Q} \quad \text{e} \quad MPP_L \equiv \frac{\partial Q}{\partial L} = -\frac{F_L}{F_Q}$$

À parte essas derivadas, podemos obter uma outra derivada parcial,

$$\frac{\partial K}{\partial L} = -\frac{F_L}{F_K}$$

da equação $F(Q, K, L) = 0$. Qual é o significado econômico de $\partial K / \partial L$? O sinal de parcial implica que outra variável, Q , está sendo mantida constante; decorre que as variações em K e L descritas por essa derivada têm as características de variações “compensatórias” projetadas para manter o produto Q constante em um nível especificado. Portanto, elas são o tipo de variações pertinentes a movimentos *ao longo* de uma *isoquanta* de produção em cujo gráfico a variável K está no eixo vertical e a variável L no eixo horizontal. Aliás, a derivada $\partial K / \partial L$ é a medida da inclinação dessa isoquanta, que é negativa no caso normal. O valor absoluto de $\partial K / \partial L$, por outro lado, é a medida da *taxa marginal de substituição técnica* entre os dois insumos.

Extensão para o caso de equações simultâneas

O teorema da função implícita também tem uma versão mais geral e poderosa, que trata das condições sob as quais um conjunto de equações simultâneas

$$\begin{aligned} F^1(y_1, \dots, y_n; x_1, \dots, x_m) &= 0 \\ F^2(y_1, \dots, y_n; x_1, \dots, x_m) &= 0 \\ &\dots\dots\dots \\ F^n(y_1, \dots, y_n; x_1, \dots, x_m) &= 0 \end{aligned} \tag{8.24}$$

definirá, com certeza, um conjunto de funções implícitas[†]

$$\begin{aligned} y_1 &= f^1(x_1, \dots, x_m) \\ y_2 &= f^2(x_1, \dots, x_m) \\ &\dots\dots\dots \\ y_n &= f^n(x_1, \dots, x_m) \end{aligned} \tag{8.25}$$

A versão generalizada do teorema afirma que:

Dado o sistema de equações (8.24), (a) se todas as funções $F^1, \dots, F^n$ tiverem derivadas parciais contínuas em relação a todas as variáveis y e x , e (b) se em um ponto $(y_{10}, \dots, y_{n0}; x_{10}, \dots, x_{m0})$ que satisfaça (8.24), o seguinte determinante jacobiano for diferente de zero:

$$|J| \equiv \left| \frac{\partial(F^1, \dots, F^n)}{\partial(y_1, \dots, y_n)} \right| \equiv \begin{vmatrix} \frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \dots & \frac{\partial F^1}{\partial y_n} \\ \frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \dots & \frac{\partial F^2}{\partial y_n} \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \dots & \frac{\partial F^n}{\partial y_n} \end{vmatrix} \neq 0$$

[†] De um outro ponto de vista, essas condições servem para garantir que as n equações em (8.24) podem, *em princípio*, ser resolvidas para as n variáveis $y_1, \dots, y_n$ — mesmo que talvez não possamos obter a solução (8.25) de uma forma explícita.

então existe uma vizinhança m -dimensional de $(x_{10}, \dots, x_{m0})$, N , na qual as variáveis $y_1, \dots, y_n$ são funções das variáveis $x_1, \dots, x_m$ na forma de (8.25). Essas funções implícitas satisfazem

$$\begin{aligned} y_{10} &= f^1(x_{10}, \dots, x_{m0}) \\ &\dots\dots\dots \\ y_{n0} &= f^n(x_{10}, \dots, x_{m0}) \end{aligned}$$

Elas também satisfazem (8.24) para toda m -upla $(x_1, \dots, x_m)$ na vizinhança de N —e, portanto conferem a (8.24) o status de um conjunto de *identidades* no que concerne a essa vizinhança. Além disso, as funções implícitas $f^1, \dots, f^n$ são contínuas e têm derivadas parciais contínuas em relação a todas as variáveis x .

Como no caso de uma única equação, é possível obter as derivadas parciais das funções implícitas diretamente das n equações em (8.24) sem ter de resolvê-las para as variáveis y . Aproveitando o fato de que, na vizinhança de N , as equações em (8.24) têm o status de identidades, podemos tomar a diferencial total de cada uma delas e escrever $dF^j = 0$ ($j = 1, 2, \dots, n$). O resultado é um conjunto de equações envolvendo as diferenciais $dy_1, \dots, dy_n$ e $dx_1, \dots, dx_m$. Especificamente, após transpor os termos dx_i para a direita do sinal de igualdade, temos

$$\begin{aligned} \frac{\partial F^1}{\partial y_1} dy_1 + \frac{\partial F^1}{\partial y_2} dy_2 + \dots + \frac{\partial F^1}{\partial y_n} dy_n &= - \left(\frac{\partial F^1}{\partial x_1} dx_1 + \dots + \frac{\partial F^1}{\partial x_m} dx_m \right) \\ \frac{\partial F^2}{\partial y_1} dy_1 + \frac{\partial F^2}{\partial y_2} dy_2 + \dots + \frac{\partial F^2}{\partial y_n} dy_n &= - \left(\frac{\partial F^2}{\partial x_1} dx_1 + \dots + \frac{\partial F^2}{\partial x_m} dx_m \right) \\ &\dots\dots\dots \\ \frac{\partial F^n}{\partial y_1} dy_1 + \frac{\partial F^n}{\partial y_2} dy_2 + \dots + \frac{\partial F^n}{\partial y_n} dy_n &= - \left(\frac{\partial F^n}{\partial x_1} dx_1 + \dots + \frac{\partial F^n}{\partial x_m} dx_m \right) \end{aligned} \quad (8.26)$$

Além disso, de (8.25), podemos escrever as diferenciais das variáveis y_j como

$$\begin{aligned} dy_1 &= \frac{\partial y_1}{\partial x_1} dx_1 + \frac{\partial y_1}{\partial x_2} dx_2 + \dots + \frac{\partial y_1}{\partial x_m} dx_m \\ dy_2 &= \frac{\partial y_2}{\partial x_1} dx_1 + \frac{\partial y_2}{\partial x_2} dx_2 + \dots + \frac{\partial y_2}{\partial x_m} dx_m \\ &\dots\dots\dots \\ dy_n &= \frac{\partial y_n}{\partial x_1} dx_1 + \frac{\partial y_n}{\partial x_2} dx_2 + \dots + \frac{\partial y_n}{\partial x_m} dx_m \end{aligned} \quad (8.27)$$

e elas podem ser utilizadas para eliminar as expressões dy_j em (8.26). Mas, uma vez que o resultado da substituição seria impraticavelmente confuso, vamos simplificar as coisas considerando apenas o que acontece quando somente x_1 varia enquanto todas as outras variáveis $x_2, \dots, x_m$ permanecem constantes. Deixando $dx_1 \neq 0$, mas estabelecendo $dx_2 = \dots = dx_m = 0$ in (8.26) e (8.27), então substituindo (8.27) em (8.26) e dividindo tudo por $dx_1 \neq 0$, obtemos o sistema de equações

$$\begin{aligned} \frac{\partial F^1}{\partial y_1} \left(\frac{\partial y_1}{\partial x_1} \right) + \frac{\partial F^1}{\partial y_2} \left(\frac{\partial y_2}{\partial x_1} \right) + \dots + \frac{\partial F^1}{\partial y_n} \left(\frac{\partial y_n}{\partial x_1} \right) &= - \frac{\partial F^1}{\partial x_1} \\ \frac{\partial F^2}{\partial y_1} \left(\frac{\partial y_1}{\partial x_1} \right) + \frac{\partial F^2}{\partial y_2} \left(\frac{\partial y_2}{\partial x_1} \right) + \dots + \frac{\partial F^2}{\partial y_n} \left(\frac{\partial y_n}{\partial x_1} \right) &= - \frac{\partial F^2}{\partial x_1} \\ &\dots\dots\dots \\ \frac{\partial F^n}{\partial y_1} \left(\frac{\partial y_1}{\partial x_1} \right) + \frac{\partial F^n}{\partial y_2} \left(\frac{\partial y_2}{\partial x_1} \right) + \dots + \frac{\partial F^n}{\partial y_n} \left(\frac{\partial y_n}{\partial x_1} \right) &= - \frac{\partial F^n}{\partial x_1} \end{aligned} \quad (8.28)$$

Mesmo esse resultado – para o caso em que somente x_1 varia – parece de formidável complexidade, por estar cheio de derivadas. Mas, na verdade, sua estrutura é bem fácil de compreender, uma vez que saibamos distinguir entre os dois tipos de derivadas que aparecem em (8.28). Um tipo, que apresentamos entre parênteses para distinguir melhor visualmente, consiste nas derivadas parciais das funções implícitas em relação a x_1 que estamos procurando. Por conseguinte, elas devem ser consideradas as “variáveis” a serem calculadas em (8.28). O outro tipo, por sua vez, consiste nas derivadas parciais das funções F^j dadas em (8.24). Visto que todas elas assumiriam valores específicos quando avaliadas no ponto $(y_{10}, \dots, y_{n0}; x_{10}, \dots, x_{m0})$ – o ponto ao redor do qual as funções implícitas são definidas – elas aparecem aqui não como funções derivadas, mas como valores de derivadas. Como tal, podem ser tratadas como constantes dadas. Esse fato faz de (8.28) um *sistema linear de equações*, com uma estrutura semelhante a (4.1). O interessante é que tal sistema linear surgiu durante o processo de análise de um problema que não é, em si, necessariamente linear, pois nenhuma restrição de linearidade foi imposta ao sistema de equações (8.24). Assim, temos aqui uma ilustração de como a álgebra linear pode estar presente mesmo em problemas não-lineares.

Na qualidade de um sistema linear de equações, (8.28) pode ser escrito em notação matricial como

$$\begin{bmatrix} \frac{\partial F^1}{\partial y_1} & \frac{\partial F^1}{\partial y_2} & \dots & \frac{\partial F^1}{\partial y_n} \\ \frac{\partial F^2}{\partial y_1} & \frac{\partial F^2}{\partial y_2} & \dots & \frac{\partial F^2}{\partial y_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial F^n}{\partial y_1} & \frac{\partial F^n}{\partial y_2} & \dots & \frac{\partial F^n}{\partial y_n} \end{bmatrix} \begin{bmatrix} \left(\frac{\partial y_1}{\partial x_1} \right) \\ \left(\frac{\partial y_2}{\partial x_1} \right) \\ \vdots \\ \left(\frac{\partial y_n}{\partial x_1} \right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial F^1}{\partial x_1} \\ -\frac{\partial F^2}{\partial x_1} \\ \vdots \\ -\frac{\partial F^n}{\partial x_1} \end{bmatrix} \quad (8.28')$$

Visto que o determinante da matriz de coeficientes em (8.28') nada mais é do que o determinante jacobiano particular $|J|$ que sabemos ser diferente de zero sob as condições do teorema da função implícita, e, posto que o sistema deve ser não-homogêneo (por quê?), deve haver uma única solução não-trivial para (8.28'). Pela regra de Cramer, essa solução pode ser expressa analiticamente da seguinte maneira:

$$\left(\frac{\partial y_j}{\partial x_1} \right) = \frac{|J_j|}{|J|} \quad (j = 1, 2, \dots, n) \quad [\text{veja (5.18)}] \quad (8.29)$$

Por uma adaptação adequada desse procedimento, as derivadas parciais das funções implícitas em relação às outras variáveis, $x_2, \dots, x_m$, também podem ser obtidas. Um aspecto muito bom desse procedimento é que, cada vez que permitirmos que uma determinada variável x_i varie, poderemos obter, de uma vez só, as derivadas parciais de todas as funções implícitas $f^1, \dots, f^n$ em relação àquela variável x_i em particular.

De modo semelhante, para a regra da função implícita (8.23) no caso de uma única equação, o procedimento que acabamos de descrever requer somente a utilização das derivadas parciais das funções F – avaliadas no ponto $(y_{10}, \dots, y_{n0}; x_{10}, \dots, x_{m0})$ – no cálculo das derivadas parciais das funções implícitas $f^1, \dots, f^n$. Assim, a equação matricial (8.28') e sua solução analítica (8.29) são, com efeito, um enunciado da regra da função implícita, em sua versão para equações simultâneas.

Note que a exigência de que $|J| \neq 0$ exclui um denominador zero em (8.29), exatamente como o fazia a exigência de que $F_y \neq 0$ na regra da função implícita (8.23) e (8.23'). Além disso, o papel desempenhado pela condição $|J| \neq 0$ na garantia de uma solução única (se bem que implícita) (8.25) para o sistema geral (possivelmente *não-linear*) (8.24) é muito semelhante ao papel da condição de não-singularidade $|A| \neq 0$ em um sistema *linear* $Ax = d$.

EXEMPLO 5 As seguintes três equações

$$xy - w = 0 \quad F^1 = (x, y, w; z) = 0$$

$$y - w^3 - 3z = 0 \quad F^2 = (x, y, w; z) = 0$$

$$w^3 + z^3 - 2zw = 0 \quad F^3 = (x, y, w; z) = 0$$

são satisfeitas no ponto $P: (x, y, w; z) = (\frac{1}{4}, 4, 1, 1)$. As funções f^i obviamente possuem derivadas contínuas. Assim, se o jacobiano $|J|$ for diferente de zero no ponto P , podemos usar o teorema da função implícita para obter a derivada estática comparativa $(\partial x / \partial z)$.

Para fazer isso, podemos primeiramente tomar a diferencial total do sistema:

$$y \, dx + x \, dy - dw = 0$$

$$dy - 3w^2 \, dw - 3 \, dz = 0$$

$$(3w^2 - 2z) \, dw + (3z^2 - 2w) \, dz = 0$$

Passando a diferencial exógena (e seus coeficientes) para o lado direito e escrevendo em forma matricial, obtemos:

$$\begin{bmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dw \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 2w - 3z^2 \end{bmatrix} dz$$

onde a matriz de coeficientes do lado esquerdo é o jacobiano

$$|J| = \begin{bmatrix} F_x^1 & F_y^1 & F_w^1 \\ F_x^2 & F_y^2 & F_w^2 \\ F_x^3 & F_y^3 & F_w^3 \end{bmatrix} = \begin{bmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{bmatrix} = y(3w^2 - 2z)$$

No ponto P , o determinante jacobiano $|J| = 4 (\neq 0)$. Portanto, a regra da função implícita se aplica e

$$\begin{bmatrix} y & x & -1 \\ 0 & 1 & -3w^2 \\ 0 & 0 & (3w^2 - 2z) \end{bmatrix} \begin{bmatrix} \left(\frac{\partial x}{\partial z} \right) \\ \left(\frac{\partial y}{\partial z} \right) \\ \left(\frac{\partial w}{\partial z} \right) \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 2w - 3z^2 \end{bmatrix}$$

Usando a regra de Cramer para encontrar uma expressão para $(\partial x / \partial z)$, obtemos:

$$\left(\frac{\partial x}{\partial z} \right) = \frac{\begin{vmatrix} 0 & x & -1 \\ 3 & 1 & -3w^2 \\ 2w - 3z^2 & 0 & (3w^2 - 2z) \end{vmatrix}}{|J|} = \frac{\begin{vmatrix} 0 & \frac{1}{4} & -1 \\ 3 & 1 & -3 \\ -1 & 0 & 1 \end{vmatrix}}{4}$$

$$= 0 + (-3) \frac{\begin{vmatrix} \frac{1}{4} & -1 \\ 0 & 1 \end{vmatrix}}{4} + (-1) \frac{\begin{vmatrix} \frac{1}{4} & -1 \\ 1 & -3 \end{vmatrix}}{4}$$

$$= \frac{-3}{16} + \frac{-1}{16}$$

$$= -\frac{1}{4}$$

EXEMPLO 6 Seja o modelo de renda nacional (7.17) reescrito na forma

$$\begin{aligned} Y - C - I_0 - G_0 &= 0 \\ C - \alpha - \beta(Y - T) &= 0 \\ T - \gamma - \delta Y &= 0 \end{aligned} \quad (8.30)$$

Se tomarmos as variáveis endógenas (Y, C, T) como (y_1, y_2, y_3) , e tomarmos as variáveis exógenas e parâmetros $(I_0, G_0, \alpha, \beta, \gamma, \delta)$ como $(x_1, x_2, \dots, x_6)$, então a expressão do lado esquerdo de cada equação pode ser considerada como uma função F específica, na forma de $F^i(Y, C, T; I_0, G_0, \alpha, \beta, \gamma, \delta)$. Assim, (8.30) é um caso específico de (8.24), com $n = 3$ e $m = 6$. Visto que as funções F^1, F^2 e F^3 têm derivadas parciais contínuas e visto que o determinante jacobiano relevante (aquele que envolve somente as variáveis endógenas),

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial Y} & \frac{\partial F^1}{\partial C} & \frac{\partial F^1}{\partial T} \\ \frac{\partial F^2}{\partial Y} & \frac{\partial F^2}{\partial C} & \frac{\partial F^2}{\partial T} \\ \frac{\partial F^3}{\partial Y} & \frac{\partial F^3}{\partial C} & \frac{\partial F^3}{\partial T} \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{vmatrix} = 1 - \beta + \beta\delta \quad (8.31)$$

seja sempre diferente de zero (e ambos, β e δ sejam restritos a ser frações positivas), podemos considerar Y, C e T como funções implícitas de $(I_0, G_0, \alpha, \beta, \gamma, \delta)$ em qualquer ponto e ao redor de qualquer ponto que satisfaça (8.30). Mas um ponto que satisfaça (8.30) seria uma solução de equilíbrio relacionada a Y^*, C^* e T^* . Por conseguinte, o teorema da função implícita nos diz que é justificável escrever

$$Y^* = f^1(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

$$C^* = f^2(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

$$T^* = f^3(I_0, G_0, \alpha, \beta, \gamma, \delta)$$

para indicar que os valores de equilíbrio das variáveis endógenas são funções implícitas das variáveis exógenas e dos parâmetros.

As derivadas parciais das funções implícitas, tais como $\partial Y^* / \partial I_0$ e $\partial Y^* / \partial G_0$, têm as características de derivadas estáticas comparativas. Para encontrá-las, precisamos apenas das derivadas parciais das funções F , avaliadas no estado de equilíbrio do modelo. Além disso, visto que $n = 3$, três delas podem ser encontradas em uma única operação. Agora suponha que mantenhamos fixas todas as variáveis exógenas e parâmetros, exceto G_0 . Então, adaptando o resultado em (8.28'), podemos escrever a equação

$$\begin{bmatrix} 1 & -1 & 0 \\ -\beta & 1 & \beta \\ -\delta & 0 & 1 \end{bmatrix} \begin{bmatrix} \partial Y^* / \partial G_0 \\ \partial C^* / \partial G_0 \\ \partial T^* / \partial G_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

da qual podem ser calculadas três derivadas estáticas comparativas (todas em relação a G_0). A primeira, que representa o multiplicador da despesa do governo, resultará, por exemplo, em

$$\frac{\partial Y^*}{\partial G_0} = \frac{\begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{1}{1 - \beta + \beta\delta} \quad [\text{por (8.31)}]$$

Isso, é claro, nada mais é do que o resultado obtido anteriormente em (7.19). Note, contudo, que, na presente abordagem, trabalhamos somente com funções implícitas e evitamos completamente a etapa da resolução do sistema (8.30) explicitamente para Y^*, C^* e T^* . É essa característica particular do método que agora nos habilitará a encarar a estática comparativa de modelos de funções gerais que, por sua própria natureza, não podem dar como resultado nenhuma solução explícita.

EXERCÍCIO 8.5

- Para cada $F(x, y) = 0$, encontre dy/dx para cada uma das seguintes:
 - $y - 6x + 7 = 0$
 - $3y + 12x + 17 = 0$
 - $x^2 + 6x - 13 - y = 0$
- Para cada $F(x, y) = 0$, use a regra da função implícita para encontrar dy/dx :
 - $F(x, y) = 3x^2 + 2xy + 4y^3 = 0$
 - $F(x, y) = 12x^5 - 2y = 0$
 - $F(x, y) = 7x^2 + 2xy^2 + 9y^4 = 0$
 - $F(x, y) = 6x^3 - 3y = 0$
- Para cada $F(x, y, z) = 0$, use a regra da função implícita para encontrar $\partial y/\partial x$ e $\partial y/\partial z$:
 - $F(x, y, z) = x^2y^3 + z^2 + xyz = 0$
 - $F(x, y, z) = x^3z^2 + y^3 + 4xyz = 0$
 - $F(x, y, z) = 3x^2y^3 + xz^2y^2 + y^3zx^4 + y^2z = 0$
- Admitindo que a equação $F(U, x_1, x_2, \dots, x_n) = 0$ define implicitamente a função utilidade $U = f(x_1, x_2, \dots, x_n)$:
 - Calcule as expressões para $\partial U/\partial x_2$, $\partial U/\partial x_n$, $\partial x_3/\partial x_2$ e $\partial x_4/\partial x_n$.
 - Interprete seus respectivos significados econômicos.
- Para cada uma das equações dadas $F(y, x) = 0$, uma função implícita $y = f(x)$ é definida ao redor do ponto $(y = 3, x = 1)$?
 - $x^3 - 2x^2y + 3xy^2 - 22 = 0$
 - $2x^2 + 4xy - y^4 + 67 = 0$
 Se sua resposta for afirmativa, calcule dy/dx pela regra da função implícita e avalie-a no ponto citado.
- Dada $x^2 + 3xy + 2yz + y^2 + z^2 - 11 = 0$, uma função implícita $z = f(x, y)$ é definida ao redor do ponto $(x = 1, y = 2, z = 0)$? Se for, calcule $\partial z/\partial x$ e $\partial z/\partial y$ pela regra da função implícita e avalie-as naquele ponto.
- Considerando a equação $F(y, x) = (x - y)^3 = 0$ em uma vizinhança ao redor do ponto de origem, demonstre que as condições citadas no teorema da função implícita *não* têm as características de condições *necessárias*.
- Se a equação $F(x, y, z) = 0$ define implicitamente cada uma das três variáveis como uma função das outras duas variáveis, e se todas as derivadas em questão existem, calcule o valor de $\frac{\partial z}{\partial x} \frac{\partial x}{\partial y} \frac{\partial y}{\partial z}$.
- Justifique a afirmativa apresentada no texto de que o sistema de equações (8.28') deve ser não-homogêneo.
- A partir do modelo de renda nacional (8.30), encontre o multiplicador não referente ao imposto sobre a renda pela regra da função implícita. Compare seus resultados com (7.20).

8.6 Estática comparativa de modelos de funções gerais

Quando consideramos pela primeira vez o problema da análise estática comparativa no Capítulo 7, tratamos do caso em que os valores de equilíbrio das variáveis endógenas do modelo são expressos explicitamente em termos das variáveis exógenas e parâmetros. Naquele caso, bastava a técnica de simples diferenciação parcial. Quando um modelo contém funções expressas na forma geral, entretanto, aquela técnica torna-se inaplicável por causa da indisponibilidade de soluções explícitas. Em seu lugar, deve ser empregada uma nova técnica, que faça uso de tais conceitos como diferenciais totais, derivadas totais, bem como do teorema da função implícita e da regra da função implícita. Isso será ilustrado primeiramente com um modelo de mercado e, em seguida, passaremos para modelos de renda nacional.

Modelo de mercado

Considere um mercado de uma única mercadoria, no qual a quantidade demandada Q_d é uma função não somente do preço P , mas também de uma renda exógena determinada, Y_0 . A quantidade oferecida, Q_s , por outro lado, é uma função apenas do preço. Se essas funções não forem dadas em formas específicas, nosso modelo pode ser escrito, de forma geral, como:

$$\begin{aligned} Q_d &= Q_s \\ Q_d &= D(P, Y_0) \quad (\partial D/\partial P < 0; \partial D/\partial Y_0 > 0) \\ Q_s &= S(P) \quad (dS/dP > 0) \end{aligned} \quad (8.32)$$

Admite-se que ambas as funções D e S possuem derivadas contínuas ou, em outras palavras, seus gráficos são suaves. Além disso, de modo a assegurar relevância econômica, impusemos restrições definidas sobre os sinais dessas derivadas. Pela restrição $dS/dP > 0$, a função oferta é definida como estritamente crescente, embora se permita que ela seja linear ou não-linear. De modo semelhante, pelas restrições impostas às duas derivadas parciais da função demanda, indicamos que ela é uma função estritamente decrescente do preço, mas uma função estritamente crescente da renda. Por questão de simplicidade de notação, as restrições de sinais impostas às derivadas de uma função às vezes são indicadas por sinais + ou – colocados diretamente embaixo das variáveis independentes. Assim, as funções D e S em (8.32) podem ser apresentadas alternativamente como

$$Q_d = D(P, Y_0) \quad Q_s = S(P)$$

$\begin{matrix} - & + \\ + & + \end{matrix}$

Essas restrições servem para confinar nossa análise ao caso “normal” que esperamos encontrar.

Ao desenhar o tipo usual de curva bidimensional de demanda, admite-se que o nível de renda é mantido fixo. Quando a renda variar, ela afetará um dado equilíbrio por causar um deslocamento da curva de demanda. De modo semelhante, em (8.32), Y_0 pode causar uma variação desequilibradora por toda a função demanda. Aqui, Y_0 é a única variável exógena ou parâmetro; assim, a análise estática comparativa desse modelo se referirá exclusivamente ao modo como uma variação em Y_0 afetará a posição de equilíbrio do modelo.

A posição de equilíbrio do mercado é definida pela condição de equilíbrio $Q_d = Q_s$, que, com substituição e rearranjo, pode ser expressa por:

$$D(P, Y_0) - S(P) = 0 \quad (8.33)$$

Mesmo que essa equação não possa ser resolvida explicitamente para o preço de equilíbrio P^* , admitiremos que realmente existe um equilíbrio estático – pois, caso contrário, não teria nenhum sentido nem mesmo levantar a questão da estática comparativa. De nossa experiência com modelos de funções específicas, aprendemos a esperar que P^* seja uma função da variável exógena Y_0 :

$$P^* = P^*(Y_0) \quad (8.34)$$

Mas agora podemos proporcionar um fundamento rigoroso para essa expectativa apelando para o teorema da função implícita. Considerando que (8.33) está na forma de $F(P, Y_0) = 0$, satisfazer as condições do teorema da função implícita garantirá que cada valor de Y_0 resulte em um valor único de P^* na vizinhança de um ponto que satisfaça (8.33), isto é, na vizinhança de uma solução de equilíbrio (inicial ou “velha”). Nesse caso, podemos de fato escrever a função implícita $P^* = P^*(Y_0)$ e discutir sua derivada, dP^*/dY_0 – aquela derivada estática comparativa que desejamos – e que sabemos que existe. Portanto, vamos verificar essas condições. Em primeiro lugar, a função $F(P, Y_0)$ realmente possui derivadas contínuas; isso porque, por premissa, seus dois componentes aditivos $D(P, Y_0)$ e $S(P)$ têm derivadas contínuas. Em segundo lugar, a derivada parcial de F em relação a P , a saber, $F_P = \partial D/\partial P - dS/dP$, é negativa e, por conseguinte, diferente de zero, não importando onde ela é avaliada. Assim, o teorema da função implícita se aplica e (8.34) é, de fato, legítima.

De acordo com o mesmo teorema, a condição de equilíbrio (8.33) agora pode ser tomada como uma identidade em alguma vizinhança da solução de equilíbrio. Por consequência, podemos escrever a identidade de equilíbrio

$$\underbrace{D(P^*, Y_0) - S(P^*)}_{F(P^*, Y_0)} \equiv 0 \quad [\text{Excesso de demanda} \equiv 0 \text{ no equilíbrio}] \quad (8.35)$$

Então, é preciso apenas uma aplicação direta da regra da função implícita para produzir a derivada estática comparativa, dP^*/dY_0 . Por questão de clareza visual, daqui em diante as derivadas estáticas comparativas serão apresentadas entre parênteses para distingui-las das expressões normais de derivadas que são meros componentes da especificação do modelo. O resultado da regra da função implícita é

$$\left(\frac{dP^*}{dY_0} \right) = - \frac{\partial F / \partial Y_0}{\partial F / \partial P^*} = - \frac{\partial D / \partial Y_0}{\partial D / \partial P^* - \partial S / \partial P^*} > 0 \quad (8.36)$$

Nesse resultado, a expressão $\partial D / \partial P^*$ refere-se à derivada $\partial D / \partial P$ avaliada no equilíbrio inicial, isto é, em $P = P^*$; uma interpretação semelhante atribui-se a $\partial S / \partial P^*$. De fato, $\partial D / \partial Y_0$ deve ser avaliada também no ponto de equilíbrio. Em virtude das especificações de sinais em (8.32), (dP^*/dY_0) é invariavelmente positiva. Assim, nossa conclusão *qualitativa* é que um aumento (redução) no nível de renda sempre resultará em um aumento (redução) no preço de equilíbrio. Se forem conhecidos os valores que as derivadas das funções de demanda e oferta assumem no equilíbrio inicial, é claro que (8.36) também resultará em uma conclusão *quantitativa*.

Essa discussão de ajuste do mercado diz respeito ao efeito de uma variação de Y_0 sobre P^* . Também é possível descobrir o efeito sobre a quantidade de equilíbrio $Q^* (= Q_d^* = Q_s^*)$? A resposta é sim. Visto que, no estado de equilíbrio, temos $Q^* = S(P^*)$ e, posto que $P^* = P^*(Y_0)$, podemos aplicar a regra da cadeia para obter a derivada

$$\left(\frac{dQ^*}{dY_0} \right) = \frac{dS}{dP^*} \left(\frac{dP^*}{dY_0} \right) > 0 \quad \left[\text{onde } \frac{dS}{dP^*} > 0 \right] \quad (8.37)$$

Assim, a quantidade de equilíbrio também está relacionada positivamente com Y_0 nesse modelo. Mais uma vez, (8.37) pode nos dar uma conclusão quantitativa se forem conhecidos os valores que as várias derivadas assumem no equilíbrio.

Os resultados em (8.36) e (8.37), que esgotam o conteúdo de estática comparativa do modelo (já que este último contém somente uma variável exógena e duas variáveis endógenas), não são surpreendentes. Na verdade, o que eles transmitem é nada mais do que a proposição de que um deslocamento da curva de demanda para cima resultará em um preço de equilíbrio mais alto, bem como em uma quantidade de equilíbrio mais alta. Isso talvez nos dê a idéia de que poderíamos chegar a essa mesma proposição em um piscar de olhos, por uma simples análise gráfica! Parece correto, mas não devemos perder de vista o caráter muito, mas muito mais geral, do procedimento analítico que temos usado aqui. A análise gráfica é, por sua própria natureza, limitada a um conjunto específico de curvas (a contraparte geométrica de um conjunto específico de funções); portanto, em termos estritos, suas conclusões são relevantes e aplicáveis somente àquele conjunto de curvas. Muito ao contrário, a formulação em (8.32), embora simplificada, abrange todo o conjunto de combinações possíveis de curvas de demanda com inclinação negativa e curvas de oferta com inclinações positivas. Isso é muitíssimo mais geral. Ademais, o procedimento analítico utilizado aqui pode manipular muitos problemas de maior complexidade que provariam estar além das capacidades da abordagem gráfica.

Abordagem de equações simultâneas

A análise do modelo (8.32) foi executada tendo como base uma única equação, a saber, (8.35). Visto que somente uma variável endógena pode ser incorporada proveitosamente em uma só equação, a inclusão de P^* significa a exclusão de Q^* . Por consequência, fomos obrigados a encon-

trar primeiro (dP^*/dY_0) e depois inferir (dQ^*/dY_0) em uma etapa subsequente. Agora mostraremos como P^* e Q^* podem ser estudados simultaneamente. Como há duas variáveis endógenas, conseqüentemente estabeleceremos um sistema de duas equações. Em primeiro lugar, sendo $Q = Q_d = Q_s$ em (8.32) e rearranjando, podemos expressar nosso modelo de mercado como

$$\begin{aligned} F^1(P, Q; Y_0) &= D(P, Y_0) - Q = 0 \\ F^2(P, Q; Y_0) &= S(P) - Q = 0 \end{aligned} \quad (8.38)$$

o que está na forma de (8.24), com $n = 2$ e $m = 1$. Mais uma vez, é interessante verificar as condições do teorema da função implícita. Primeiro, visto que se admite que ambas as funções de demanda e de oferta possuem derivadas contínuas, o mesmo deve ser suposto para as funções F^1 e F^2 . Segundo, o jacobiano de variáveis endógenas (aquele que envolve P e Q) realmente revela ser diferente de zero, independentemente de onde ele é avaliado, porque

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial P} & \frac{\partial F^1}{\partial Q} \\ \frac{\partial F^2}{\partial P} & \frac{\partial F^2}{\partial Q} \end{vmatrix} = \begin{vmatrix} \frac{\partial D}{\partial P} & -1 \\ \frac{dS}{dP} & -1 \end{vmatrix} = \frac{dS}{dP} - \frac{\partial D}{\partial P} > 0 \quad (8.39)$$

Portanto, se existir uma solução de equilíbrio (P^*, Q^*) (o que devemos admitir, para que faça sentido falar em estática comparativa), o teorema da função implícita nos diz que podemos escrever as funções implícitas

$$P^* = P^*(Y_0) \quad \text{e} \quad Q^* = Q^*(Y_0) \quad (8.40)$$

mesmo que não possamos resolvê-las explicitamente para P^* e Q^* . Sabe-se que essas funções têm derivadas contínuas. Além disso, (8.38) terá o status de um par de identidades em alguma vizinhança do estado de equilíbrio, de modo que também podemos escrever

$$\begin{aligned} D(P^*, Y_0) - Q^* &\equiv 0 \quad [\text{isto é, } F^1(P^*, Q^*; Y_0) \equiv 0] \\ S(P^*) - Q^* &\equiv 0 \quad [\text{isto é, } F^2(P^*, Q^*; Y_0) \equiv 0] \end{aligned} \quad (8.41)$$

A partir dessas, (dP^*/dY_0) e (dQ^*/dY_0) podem ser obtidas simultaneamente usando-se a regra da função implícita (8.28').

No presente contexto, com F^1 e F^2 como definidas em (8.41), e com duas variáveis endógenas P^* e Q^* e uma única variável exógena, Y_0 , a regra da função implícita assume a forma específica

$$\begin{bmatrix} \frac{\partial F^1}{\partial P^*} & \frac{\partial F^1}{\partial Q^*} \\ \frac{\partial F^2}{\partial P^*} & \frac{\partial F^2}{\partial Q^*} \end{bmatrix} = \begin{bmatrix} \left(\frac{dP^*}{dY_0} \right) \\ \left(\frac{dQ^*}{dY_0} \right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial F^1}{\partial Y_0} \\ -\frac{\partial F^2}{\partial Y_0} \end{bmatrix}$$

Note que as derivadas estáticas comparativas são escritas aqui com o símbolo d em vez de ∂ , porque há somente uma variável exógena no presente problema. Mais especificamente, a última equação pode ser expressa como

$$\begin{bmatrix} \frac{\partial D}{\partial P^*} & -1 \\ \frac{dS}{dP^*} & -1 \end{bmatrix} = \begin{bmatrix} \left(\frac{dP^*}{dY_0} \right) \\ \left(\frac{dQ^*}{dY_0} \right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial D}{\partial Y_0} \\ 0 \end{bmatrix}$$

Pela regra de Cramer e usando (8.39), achamos então a solução, que é:

$$\begin{aligned} \left(\frac{dP^*}{dY_0} \right) &= \frac{\begin{vmatrix} -\frac{\partial D}{\partial Y_0} & -1 \\ 0 & -1 \end{vmatrix}}{|J|} = \frac{\frac{\partial D}{\partial Y_0}}{|J|} \\ \left(\frac{dQ^*}{dY_0} \right) &= \frac{\begin{vmatrix} \frac{\partial D}{\partial P^*} & \frac{\partial D}{\partial Y_0} \\ \frac{dS}{dP^*} & 0 \end{vmatrix}}{|J|} = \frac{\frac{dS}{dP^*} \frac{\partial D}{\partial Y_0}}{|J|} \end{aligned} \quad (8.42)$$

onde todas as derivadas das funções demanda e oferta (incluindo as que aparecem no jacobiano) devem ser avaliadas no equilíbrio inicial. Você pode verificar que os resultados que acabamos de obter são idênticos aos obtidos anteriormente em (8.36) e (8.37), por meio da abordagem da equação única.

Em vez de aplicar a regra da função implícita diretamente, também podemos chegar ao mesmo resultado primeiramente diferenciando totalmente cada uma das identidades em (8.41) por vez, para obter um sistema linear de equações com as variáveis dP^* e dQ^*

$$\begin{aligned} \frac{\partial D}{\partial P^*} dP^* - dQ^* &= -\frac{\partial D}{\partial Y_0} dY_0 \\ \frac{dS}{dP^*} dP^* - dQ^* &= 0 \end{aligned}$$

e, então, dividindo tudo por $dY_0 \neq 0$ e interpretando cada quociente de duas diferenciais como uma derivada.

Utilização de derivadas totais

Em ambas as abordagens ilustradas anteriormente, a da equação única e a das equações simultâneas, tomamos as *diferenciais totais* de ambos os lados de uma identidade de equilíbrio e então igualamos os dois resultados para chegar à regra da função implícita. Em vez de tomar as diferenciais totais, contudo, é possível tomar, e igualar, as *derivadas totais* dos dois lados da identidade de equilíbrio em relação a uma variável exógena ou a um parâmetro em particular.

Na abordagem da equação única, por exemplo, a identidade de equilíbrio é

$$D(P^*, Y_0) - S(P^*) \equiv 0 \quad [\text{de (8.35)}]$$

$$\text{onde} \quad P^* = P^*(Y_0) \quad [\text{de (8.34)}]$$

Tomar a derivada total da identidade de equilíbrio em relação a Y_0 – o que leva em conta os efeitos indiretos, bem como os efeitos diretos de uma variação em Y_0 – nos dará, por conseguinte, a equação

$$\underbrace{\frac{\partial D}{\partial P^*} \left(\frac{dP^*}{dY_0} \right)}_{\text{(efeito indireto de } Y_0 \text{ sobre } D)}} + \underbrace{\frac{\partial D}{\partial Y_0}}_{\text{(efeito direto de } Y_0 \text{ sobre } D)}} - \underbrace{\frac{dS}{dP^*} \left(\frac{dP^*}{dY_0} \right)}_{\text{(efeito indireto de } Y_0 \text{ sobre } S)}} = 0$$

Quando essa expressão é resolvida para (dP^*/dY_0) , o resultado é idêntico ao resultado em (8.36).

Na abordagem de equações simultâneas, por outro lado, há um par de identidades de equilíbrio:

$$D(P^*, Y_0) - Q^* \equiv 0$$

$$S(P^*) - Q^* \equiv 0 \quad [\text{de (8.41)}]$$

onde $P^* = P^*(Y_0) \quad Q^* = Q^*(Y_0) \quad [\text{de (8.40)}]$

Agora fica mais difícil monitorar os vários efeitos de Y_0 , mas, com a ajuda do mapa de canal na Figura 8.7, o padrão deve ficar claro. Esse mapa de canal nos diz, por exemplo, que, ao diferenciar a função D em relação a Y_0 , devemos levar em conta o efeito indireto de Y_0 sobre D por meio de P^* , bem como o efeito direto de Y_0 (seta em curva). Por outro lado, ao diferenciar a função S em relação a Y_0 , há somente o efeito indireto (por meio de P^*) a ser levado em conta. Assim, o resultado da diferenciação total das duas identidades em relação a Y_0 é, após rearranjo, o seguinte par de equações:

$$\frac{\partial D}{\partial P^*} \left(\frac{dP^*}{dY_0} \right) - \left(\frac{dQ^*}{dY_0} \right) = - \frac{\partial D}{\partial Y_0}$$

$$\frac{dS}{dP^*} \left(\frac{dP^*}{dY_0} \right) - \left(\frac{dQ^*}{dY_0} \right) = 0$$

Elas são, é claro, idênticas às equações obtidas pelo método da diferencial total, e levam novamente às derivadas estáticas comparativas em (8.42).

Modelo da renda nacional (IS-LM)

Uma aplicação típica do teorema da função implícita é uma forma funcional geral do modelo IS-LM.[†] Nesse modelo macroeconômico, o equilíbrio é caracterizado por uma combinação de nível de renda e taxas de juros que, simultaneamente, produzem equilíbrio no mercado de bens e no mercado monetário.

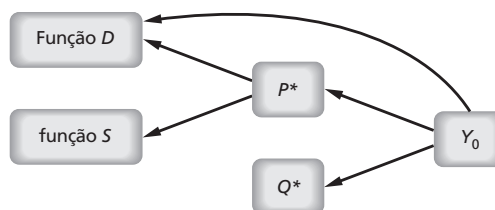
Um mercado de bens é descrito pelo seguinte conjunto de equações:

$$Y = C + I + G \quad C = C(Y - T) \quad G = G_0$$

$$I = I(r) \quad T = T(Y)$$

Y é o nível do produto interno bruto (PIB), ou renda nacional. Nessa forma do modelo, Y também pode ser imaginado como oferta agregada. C , I , G e T são consumo, investimento, despesas do governo e impostos, respectivamente.

FIGURA 8.7



[†] IS quer dizer "investimento igual à poupança" (investment equals savings) e LM quer dizer "preferência pela liquidez igual à oferta de moeda" (liquidity preference equals money supply).

1. Admite-se que o consumo é uma função estritamente crescente da renda disponível ($Y - T$). Se denotarmos a renda disponível como $Y^d = Y - T$, então a função consumo pode ser expressa como

$$C = C(Y^d)$$

onde $dC/dY^d = C'(Y^d)$ é a propensão marginal a consumir ($0 < C'(Y^d) < 1$).

2. Admite-se que a despesa de investimento é uma função estritamente decrescente da taxa de juros, r :

$$\frac{dI}{dr} = I'(r) < 0$$

3. O setor público é descrito por duas variáveis: despesas do governo (G) e impostos (T). Tipicamente, admite-se que as despesas do governo são exógenas (estabelecidas por políticas), ao passo que os impostos são considerados uma função crescente da renda. $\frac{dT}{dY} = T'(Y)$ é a taxa marginal de juros ($0 < T'(Y) < 1$).

Se substituirmos as funções C , I , G na primeira equação $Y = C + I + G$, obtemos:

$$Y = C(Y - T(Y)) + I(r) + G_0 \quad (\text{curva IS})$$

que nos dá uma única equação com duas variáveis endógenas: Y e r . Essa equação nos dá todas as combinações entre Y e r que produzem equilíbrio no mercado de bens. Essa equação define implicitamente a curva IS.

Inclinação da curva IS

Se reescrevermos a equação IS, que tem as características de uma identidade de equilíbrio,

$$Y - C(Y^d) - I(r) - G_0 \equiv 0$$

então a diferencial total relativa a Y e r é

$$dY - C'(Y^d)[1 - T'(Y)] dY - I'(r) dr = 0$$

Nota:

$$\frac{dY^d}{dY} = 1 - T'(Y)$$

Podemos rearranjar os termos dY e dr para obter uma expressão para a inclinação da curva IS:

$$\frac{dr}{dY} = \frac{1 - C'(Y^d)[1 - T'(Y)]}{I'(r)} < 0$$

Dadas as restrições impostas às derivadas de C , I e T , podemos verificar facilmente que a inclinação da curva IS é negativa.

O mercado monetário pode ser descrito pelas três equações seguintes:

$$M^d = L(Y, r) \quad [\text{demanda de moeda}] \quad \text{onde} \quad L_Y > 0 \quad \text{e} \quad L_r < 0$$

$$M^s = M_0^S \quad [\text{oferta de moeda}]$$

onde se admite que a oferta de moeda é determinada exogenamente pela autoridade monetária central e

$$M^d = M^s \quad [\text{condição de equilíbrio}]$$

Substituindo as duas primeiras equações na terceira, obtemos uma expressão que define implicitamente a curva LM, que, mais uma vez, tem as características de uma identidade de equilíbrio.

$$L(Y, r) \equiv M_0^S$$

Inclinação da curva LM

Visto que esta é uma identidade de equilíbrio, podemos tomar a diferencial total em relação às duas variáveis endógenas, Y e r :

$$L_Y dY + L_r dr = 0$$

que podem ser rearranjadas para nos dar uma expressão para a inclinação da curva LM

$$\frac{dr}{dY} = -\frac{L_Y}{L_r} > 0$$

Visto que $L_Y > 0$ e $L_r < 0$, vemos que a inclinação da curva LM é positiva.

O estado de equilíbrio macroeconômico simultâneo de ambos os mercados – de bens e monetário – pode ser descrito pelo seguinte sistema de equações:

$$Y \equiv C(Y^d) + I(r) + G_0$$

$$L(Y, r) \equiv M_0^S$$

que define implicitamente as duas variáveis endógenas, Y e r , como funções das variáveis exógenas, G_0 e M_0^S . Tomando a diferencial total do sistema, obtemos:

$$dY - C'(Y^d)[1 - T'(Y)] dY - I'(r) dr = dG_0$$

$$L_Y dY + L_r dr = dM_0^S$$

ou, em forma matricial,

$$\begin{bmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} dY \\ dr \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^S \end{bmatrix}$$

O determinante jacobiano é

$$\begin{aligned} |J| &= \begin{vmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{vmatrix} \\ &= \{1 - C'(Y^d)[1 - T'(Y)]\} L_r + L_Y I'(r) < 0 \end{aligned}$$

Visto que $|J| \neq 0$, esse sistema satisfaz as condições do teorema da função implícita, e as funções implícitas

$$Y^* = Y^*(G_0, M_0^S)$$

e

$$r^* = r^*(G_0, M_0^S)$$

podem ser escritas mesmo que não possamos resolvê-las explicitamente para Y^* e r^* . Embora não possamos resolver explicitamente para Y^* e r^* , podemos fazer exercícios de estática compa-

rativa para determinar os efeitos da variação de uma única das variáveis exógenas (G_0, M_0^S) sobre os valores de equilíbrio de Y^* e r^* . Considere as derivadas estáticas comparativas $\partial Y^*/\partial G_0$ e $\partial r^*/\partial G_0$, que, derivaremos aplicando o teorema da função implícita a nosso sistema de diferenciais totais em forma matricial

$$\begin{bmatrix} 1 - C'(Y^d)[1 - T'(Y)] & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} dY \\ dr \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^S \end{bmatrix}$$

Primeiro, fazemos $dM_0^S = 0$ e dividimos ambos os lados por dG_0 .

$$\begin{bmatrix} 1 - C' \cdot (1 - T') & -I'(r) \\ L_Y & L_r \end{bmatrix} \begin{bmatrix} \frac{dY^*}{dG_0} \\ \frac{dr^*}{dG_0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Usando a regra de Cramer, obtemos:

$$\frac{dY^*}{dG_0} = \frac{\begin{vmatrix} 1 & -I' \\ 0 & L_r \end{vmatrix}}{|J|} = \frac{L_r}{|J|} = \frac{\Theta}{\Theta} > 0$$

e

$$\frac{dr^*}{dG_0} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') & 1 \\ L_Y & 0 \end{vmatrix}}{|J|} = \frac{-L_Y}{|J|} = \frac{\Theta}{\Theta} > 0$$

Segundo o teorema da função implícita, essas razões de diferenciais, dY^*/dG_0 e dr^*/dG_0 , podem ser interpretadas como derivadas parciais,

$$\frac{\partial Y^*(G_0, M_0^S)}{\partial G_0} \quad \text{e} \quad \frac{\partial r^*(G_0, M_0^S)}{\partial G_0}$$

que são as derivadas estáticas comparativas que desejávamos.

Ampliando o modelo: uma economia aberta

Uma propriedade que os economistas buscam em modelos é a sua robustez; a capacidade de o modelo ser aplicado a diferentes cenários. Neste ponto, ampliaremos o modelo básico para incorporar o setor externo.

1. *Exportações líquidas.* Denote por X as exportações, por M as importações e por E a taxa de câmbio (medida como o preço interno da moeda estrangeira). Exportações são uma função crescente da taxa da câmbio.

$$X = X(E) \quad \text{onde} \quad X'(E) > 0$$

Importações são uma função decrescente da taxa de câmbio, mas uma função crescente da renda.

$$M = M(Y, E) \quad \text{onde} \quad M_Y > 0, M_E < 0$$

2. *Fluxos de capital.* O fluxo líquido de capital que entra em um país é uma função da taxa interna de juros r e também da taxa de juros mundial r_w . Vamos denotar por K o fluxo líquido de entrada de capital, tal que

$$K = K(r, r_w) \quad \text{onde} \quad K_r > 0, K_{r_w} < 0$$

3. *Balança de pagamentos.* Os fluxos de entrada e de saída de moeda estrangeira em um país são tipicamente separados em duas contas: a conta-corrente (exportações líquidas de bens e serviços) e a conta de *capital* (compra de títulos estrangeiros e nacionais). Juntas, as duas contas compõem a balança de pagamentos.

$$BP = \text{conta-corrente} + \text{conta de capital}$$

$$= [X(E) - M(Y, E)] + K(r, r_w)$$

Sob regime de taxas de câmbio flutuantes, a taxa de câmbio é ajustada para manter a balança de pagamentos igual a zero. Balança de pagamentos igual a zero equivale a dizer que a oferta de moeda estrangeira é igual à demanda de moeda estrangeira por um país.[†]

O equilíbrio na economia aberta

Equilíbrio em uma economia aberta é caracterizado por três condições: demanda agregada é igual à oferta agregada; a demanda por moeda é igual à oferta de moeda; a balança de pagamentos é igual a zero. O acréscimo do setor externo ao nosso modelo básico nos dá o seguinte sistema de três equações

$$Y = C(Y^d) + I(r) + G_0 + X(E) - M(Y, E)$$

$$L(Y, r) = M_0^s$$

$$X(E) - M(Y, E) + K(r, r_w) = 0$$

Visto que temos três equações, precisamos de três variáveis endógenas, que são Y , r e E . As variáveis exógenas agora se tornam G_0 , M_0^s e r_w . Reescrever o sistema como identidades de equilíbrio $F^1 \equiv 0$, $F^2 \equiv 0$, $F^3 \equiv 0$ nos permite achar o jacobiano:

$$Y - C(Y^d) - I(r) - G_0 - X(E) + M(Y, E) \equiv 0$$

$$L(Y, r) - M_0^s \equiv 0$$

$$X(E) - M(Y, E) + K(r, r_w) \equiv 0$$

$$|J| = \begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{vmatrix}$$

[†] Sob um regime de taxa de câmbio fixa, a balança de pagamentos não é necessariamente zero. Caso seja, quaisquer superávits ou déficits são registrados como *mudança de acordos oficiais*.

Usando a expansão de Laplace na terceira coluna, obtemos:

$$\begin{aligned}
 |J| &= (M_E - X') \begin{vmatrix} L_Y & L_r \\ -M_Y & K_r \end{vmatrix} + (X' - M_E) \begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' \\ L_Y & L_r \end{vmatrix} \\
 &= (M_E - X') (L_Y K_r + L_r M_Y) + (X' - M_E) \{ [1 - C' \cdot (1 - T') + M_Y] L_r + I' L_Y \} \\
 &= (M_E - X') \{ L_Y (K_r - I') + L_r [C' (1 - T') - 1] \}
 \end{aligned}$$

Dadas as premissas sobre os sinais das derivadas parciais e a restrição de que $0 < C' \cdot (1 - T') < 1$, podemos determinar que $|J| < 0$. Por conseguinte, podemos escrever as funções implícitas

$$Y^* = Y^*(G_0, M_0^s, r_w)$$

$$r^* = r^*(G_0, M_0^s, r_w)$$

$$E^* = E^*(G_0, M_0^s, r_w)$$

Tomar a diferencial total do sistema de equações e escrevê-la em forma matricial

$$\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{vmatrix} \begin{bmatrix} dY^* \\ dr^* \\ dE^* \end{bmatrix} = \begin{bmatrix} dG_0 \\ dM_0^s \\ -K_{r_w} dr_w \end{bmatrix}$$

nos permitirá efetuar uma série de exercícios de estática comparativa. Vamos considerar o impacto da variação na taxa mundial de juros r_w sobre os valores de equilíbrio de Y , r e E . Estabelecendo $dG_0 = dM_0^s = 0$ e dividindo ambos os lados por dr_w , temos:

$$\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & M_E - X' \\ L_Y & L_r & 0 \\ -M_Y & K_r & X' - M_E \end{vmatrix} \begin{bmatrix} \frac{dY^*}{dr_w} \\ \frac{dr^*}{dr_w} \\ \frac{dE^*}{dr_w} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -K_{r_w} \end{bmatrix}$$

Usando a regra de Cramer, obtemos as derivadas estáticas comparativas

$$\frac{\partial Y^*}{\partial r_w} = \frac{\begin{vmatrix} 0 & -I' & M_E - X' \\ 0 & L_r & 0 \\ -K_{r_w} & K_r & X' - M_E \end{vmatrix}}{|J|} = \frac{(-K_{r_w})(-L_r)(M_E - X')}{|J|} > 0$$

e

$$\frac{\partial r^*}{\partial r_w} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & 0 & M_E - X' \\ L_Y & 0 & 0 \\ -M_Y & -K_{r_w} & X' - M_E \end{vmatrix}}{|J|} = \frac{K_{r_w}(-L_Y)(M_E - X')}{|J|} > 0$$

e

$$\frac{\partial E^*}{\partial r_w} = \frac{\begin{vmatrix} 1 - C' \cdot (1 - T') + M_Y & -I' & 0 \\ L_Y & L_r & 0 \\ -M_Y & K_r & -K_{rw} \end{vmatrix}}{|J|}$$

$$= \frac{-K_{rw} \{ [1 - C' \cdot (1 - T') + M_Y] L_r + L_Y I' \}}{|J|} > 0$$

Neste ponto, você deve comparar os resultados que derivamos com os princípios macroeconômicos. Intuitivamente, um aumento na taxa mundial de juros levaria a um aumento da saída de capitais e a uma depreciação da moeda nacional. Isso, por sua vez, levará a um aumento das exportações líquidas e da renda. O aumento da renda interna causará um aumento na demanda de moeda, aplicando uma pressão de alta nas taxas de juros internas. Esse resultado é ilustrado graficamente na Figura 8.8, onde um aumento na taxa mundial de juros leva a um deslocamento da curva IS para a direita.

Resumo do procedimento

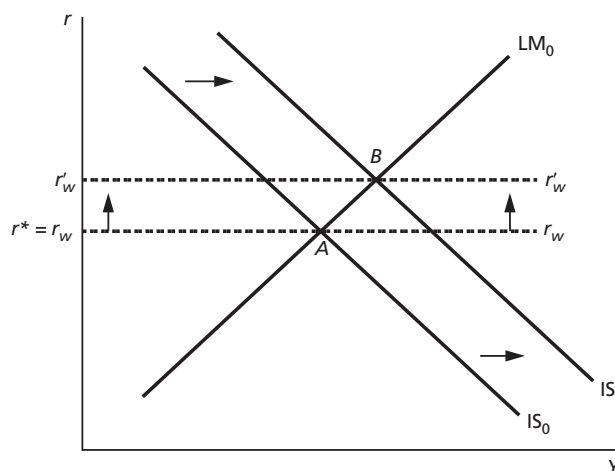
Na análise do modelo de mercado e do modelo de renda nacional de função geral, não é possível obter valores de solução explícitos das variáveis endógenas. Em vez disso, recorreremos ao teorema da função implícita que nos habilita a escrever as soluções implícitas tais como

$$P^* = P^*(Y_0) \quad \text{e} \quad r^* = r^*(G_0, M_0^s)$$

Encerramos, então, nossa busca subsequente por derivadas estáticas comparativas tais como (dP^*/dY_0) e $(\partial r^*/\partial G_0)$ por ser desprovida de sentido devido ao fato conhecido – graças ao teorema da função implícita – de que as funções P^* e r^* realmente possuem derivadas contínuas.

Para facilitar a aplicação desse teorema, adotamos uma prática padrão de escrever a condição (ou condições) de equilíbrio do modelo na forma de (8.19) ou (8.24). Então, verificamos (1) se a função (ou funções) F tem derivadas contínuas e (2) se o valor de F_y ou do determinante jacobiano de variáveis endógenas (conforme o caso) é diferente de zero no equilíbrio inicial do modelo. Todavia, contanto que as funções individuais no modelo tenham derivadas contínuas – uma premissa que é freqüentemente adotada como fato consumado em modelos de função geral – a primeira condição está automaticamente satisfeita. Portanto, na prática, basta apenas verificar o valor de F_y ou o jacobiano de variáveis endógenas. Se for diferente de zero no equilíbrio, podemos passar imediatamente à tarefa de achar derivadas estáticas comparativas.

FIGURA 8.8



A regra da função implícita é útil para tal finalidade. No caso de uma única equação, simplesmente estabeleça as variáveis endógenas iguais a seus valores de equilíbrio (por exemplo, estabeleça $P = P^*$) na condição de equilíbrio e depois aplique a regra como enunciada em (8.23) à identidade de equilíbrio resultante. No caso de equações simultâneas, devemos primeiramente estabelecer todas as variáveis endógenas iguais a seus respectivos valores de equilíbrio nas condições de equilíbrio. Então, podemos ou aplicar a regra da função implícita como ilustrada em (8.29) às identidades de equilíbrio resultantes, ou chegar ao mesmo resultado executando as várias etapas esboçadas a seguir:

1. Tome a diferencial total de cada identidade de equilíbrio por vez.
2. Selecione uma e somente uma variável exógena (digamos, X_0) como o fator exclusivo de desequilíbrio e estabeleça as diferenciais de *todas as outras* variáveis exógenas como iguais a zero. Então, divida todos os termos remanescentes em cada identidade por dX_0 e interprete cada quociente de duas diferenciais como uma derivada estática comparativa – uma derivada *parcial* se o modelo contiver duas ou mais variáveis exógenas.[†]
3. Resolva o sistema de equações resultante para as derivadas estáticas comparativas que ali aparecem e interprete suas implicações na economia. Nessa fase, se utilizarmos a regra de Cramer, podemos aproveitar a vantagem de termos verificado anteriormente a condição $|J| \neq 0$ e, por isso, na verdade, já calculamos o determinante da matriz de coeficientes do sistema de equações que está sendo resolvido agora.
4. Para analisar um outro fator de desequilíbrio (uma outra variável exógena), se houver mais um, repita as etapas 2 e 3. Surgirá um grupo diferente de derivadas estáticas comparativas no novo sistema de equações, mas a matriz de coeficientes será a mesma de antes e, assim, o valor conhecido de $|J|$ pode ser posto em uso mais uma vez.

Dado um modelo com m variáveis exógenas, serão necessárias exatamente m aplicações das etapas 1, 2 e 3 para pegar todas as derivadas estáticas comparativas que existem.

EXERCÍCIO 8.6

1. Seja a condição de equilíbrio para a renda nacional

$$S(Y) + T(Y) = I(Y) + G_0 \quad (S', T', I' > 0; S' + T' > I')$$
 onde S, Y, T, I e G representam poupança, renda nacional, impostos, investimento e despesas do governo, respectivamente. Todas as derivadas são contínuas.
 - (a) Interprete os significados econômicos das derivadas S', T' e I' .
 - (b) Verifique se as condições do teorema da função implícita foram satisfeitas. Caso positivo, escreva a identidade de equilíbrio.
 - (c) Encontre (dY^*/dG_0) e discuta suas implicações econômicas.
2. Sejam as funções demanda e oferta para uma mercadoria

$$Q_d = D(P, Y_0) \quad (D_P < 0; D_{Y_0} > 0)$$

$$Q_s = S(P, T_0) \quad (S_P > 0; S_{T_0} < 0)$$
 onde Y_0 é renda e T_0 é o imposto sobre a mercadoria. Todas as derivadas são contínuas.
 - (a) Escreva a condição de equilíbrio em uma única equação.
 - (b) Verifique se o teorema da função implícita é aplicável. Se for, escreva a identidade de equilíbrio.
 - (c) Calcule $(\partial P^*/\partial Y_0)$ e $(\partial P^*/\partial T_0)$ e discuta suas implicações econômicas.
 - (d) Usando um procedimento similar a (8.37), calcule $(\partial Q^*/\partial Y_0)$ a partir da função oferta e $(\partial Q^*/\partial T_0)$ a partir da função demanda. (Por que não usar a função demanda para a primeira e a função oferta para a última?)
3. Resolva o Problema 2 pela abordagem das equações simultâneas.

[†] Em vez de executar as etapas 1 e 2, podemos igualmente recorrer ao método da derivada total diferenciando (ambos os lados) de cada identidade de equilíbrio totalmente em relação à variável exógena escolhida. Para fazer isso, será útil usar um mapa de canal.

4. Sejam as funções demanda e oferta para uma mercadoria
- $$Q_d = D(P, t_0) \quad \text{e} \quad Q_s = Q_{s0}$$
- onde t_0 é o gosto dos consumidores pela mercadoria e onde ambas as derivadas parciais são contínuas.
- Qual é o significado dos sinais – e + embaixo das variáveis independentes P e t_0 ?
 - Escreva a condição de equilíbrio como uma única equação.
 - O teorema da função implícita é aplicável?
 - Como seria a variação do preço de equilíbrio conforme o gosto dos consumidores?
5. Considere o seguinte modelo de renda nacional (ignorando os impostos):
- $$Y - C(Y) - I(i) - G_0 = 0 \quad (0 < C' < 1; I' < 0)$$
- $$kY + L(i) - M_{s0} = 0 \quad (k = \text{constante positiva}; L' < 0)$$
- A primeira equação tem as características de uma condição de equilíbrio?
 - Qual é a quantidade total demandada por moeda neste modelo?
 - Análise a estática comparativa do modelo quando a oferta de moeda variar (política monetária) e quando a despesa do governo variar (política fiscal).
6. No Problema 5, suponha que, conquanto a demanda por moeda ainda dependa de Y como especificado, agora ela não é mais afetada pela taxa de juros.
- Como o enunciado do modelo deveria ser revisado?
 - Escreva o novo jacobiano, denomine-o $|J'|$. $|J'|$ é numericamente (em valor absoluto) maior ou menor que $|J|$?
 - A regra da função implícita ainda se aplica?
 - Calcule as novas derivadas estáticas comparativas.
 - Comparando a nova $(\partial Y^* / \partial G_0)$ com a do Problema 5, o que você pode concluir sobre a efetividade da política fiscal no novo modelo, onde Y é independente de i ?
 - Comparando a nova $(\partial Y^* / \partial M_{s0})$ com a do Problema 5, o que você pode dizer sobre a efetividade da política monetária no novo modelo?

8.7 Limitações da estática comparativa

A estática comparativa é uma área de estudo útil porque, em economia, muitas vezes estamos interessados em descobrir como uma variação desequilibradora em um parâmetro afetará o estado de equilíbrio de um modelo. Entretanto, é importante entender que, por natureza, a estática comparativa ignora o processo de ajustamento do antigo equilíbrio para o novo e também despreza o período de tempo requerido por aquele processo de ajustamento. Por conseguinte, ela deve, necessariamente, também desconsiderar a possibilidade de que, por causa da instabilidade inerente do modelo, o novo equilíbrio talvez nunca possa ser alcançado. O estudo do processo de ajustamento, por si, pertence ao campo da *economia dinâmica*. Quando chegarmos a ela, daremos particular atenção ao modo como uma variável mudará ao longo do tempo e também consideraremos explicitamente a questão da estabilidade de equilíbrio.

Contudo, o importante tópico da dinâmica tem de esperar sua vez. Enquanto isso, na Parte 4, passaremos ao estudo do problema da *otimização*, uma variedade especial extremamente importante da análise de equilíbrio com suas próprias implicações (e complicações) estáticas comparativas que a acompanham.

Parte 4

Problemas de otimização



CAPÍTULO 9

Otimização: uma variedade especial de análise de equilíbrio

Quando apresentamos o termo equilíbrio pela primeira vez no Capítulo 3, fizemos uma ampla distinção entre equilíbrio-alvo e equilíbrio que não é um alvo. No último tipo, cujo exemplo foi nosso estudo de modelos de mercado e de renda nacional, a interação entre certas forças opostas no modelo – por exemplo, as forças de demanda e de oferta nos modelos de mercado e as forças de fugas e injeções nos modelos de renda – determinam um estado de equilíbrio, se houver, no qual essas forças opostas são apenas equilibradas umas com as outras, impedindo, assim, qualquer tendência ulterior de variação. A obtenção desse tipo de equilíbrio é o resultado do equilíbrio impessoal dessas forças e não requer o esforço consciente de ninguém para atingir uma meta especificada. É verdade que os domicílios consumidores que originam as forças de demanda e as empresas que originam as forças de oferta estão, cada qual por seu lado, lutando por uma posição ótima sob as circunstâncias dadas, mas, no que diz respeito ao mercado em si, ninguém está visando a nenhum preço de equilíbrio ou a nenhuma quantidade de equilíbrio em particular (a menos, é claro, que o governo esteja tentando intervir no preço). De modo semelhante, na determinação da renda nacional, o equilíbrio impessoal entre fugas e injeções é o que cria um estado de equilíbrio, e não é necessário envolver absolutamente nenhum esforço consciente para atingir uma meta determinada (tal como uma tentativa de alterar um nível de renda indesejável por meio de políticas fiscais e monetárias).

Nesta parte do livro, contudo, voltaremos nossa atenção ao estudo do *equilíbrio-alvo*, no qual o estado de equilíbrio é definido como a posição ótima para uma dada unidade econômica (um domicílio, uma empresa comercial ou até mesmo toda uma economia) e no qual a unidade econômica em questão fará esforços deliberados para atingir aquele equilíbrio. Por consequência, nesse contexto – mas somente nesse contexto – a advertência que fizemos antes, isto é, que o equilíbrio não implica desejabilidade, torna-se irrelevante e imaterial. Nesta parte do livro, focalizaremos primordialmente as técnicas clássicas para localizar posições ótimas – as que utilizam cálculo diferencial. Aperfeiçoamentos mais modernos, conhecidos como programação matemática, serão discutidos no Capítulo 13.

9.1 Valores ótimos e valores extremos

A economia é, em essência, uma ciência de escolha. Quando um projeto econômico está para ser executado, tal como a produção de um nível especificado de produto, normalmente há inúmeros modos alternativos de realizá-lo. Entretanto, um (ou mais) desses modos alternativos será mais desejável que outros do ponto de vista de algum critério, e escolher a melhor alternativa disponível com base naquele critério especificado é a essência do problema da otimização.

O critério mais comum para escolher entre alternativas na economia é a meta de *maximizar* alguma coisa (tal como maximizar o lucro de uma empresa, a vantagem de um consumidor ou a taxa de crescimento de uma empresa ou da economia de um país) ou de *minimizar* alguma coisa (tal como minimizar o custo de produção para uma dada quantidade de produto). Em termos econômicos, podemos categorizar esses problemas de maximização e minimização sob o título geral de *otimização*, no sentido de “buscar o melhor”. Contudo, de um ponto de vista puramente matemático, os termos *máximo* e *mínimo* não trazem em si nenhuma conotação de qualificação como ótimos. Por conseguinte, o termo coletivo para máximo e mínimo como conceitos matemáticos é a designação mais neutra *extremo*, no sentido de um valor extremo.

Ao formular um problema de otimização, a primeira coisa que deve ser feita é esboçar uma *função objetivo* na qual a variável dependente representa o objeto de maximização ou minimização e na qual o conjunto de variáveis independentes indica os objetos cujas grandezas a unidade econômica em questão pode escolher com vistas à otimização. Portanto, vamos nos referir às variáveis independentes como *variáveis de escolha*.[†] A essência do processo de otimização é simplesmente achar o conjunto de valores das variáveis de escolha que nos levará ao extremo desejado da função objetivo.

Por exemplo, uma empresa comercial pode buscar maximizar o lucro, π , isto é, maximizar a diferença entre a receita total R e o custo total C . Visto que, dentro da estrutura de um dado estado de tecnologia e de uma dada demanda de mercado para o produto da empresa, R e C são, ambas, funções do nível de produção Q , decorre que π também pode ser expressa como uma função de Q :

$$\pi(Q) = R(Q) - C(Q)$$

Essa equação constitui a função objetivo relevante, na qual π é o objeto da maximização e Q é a (única) variável de escolha. Então, o problema da maximização é escolher o nível de Q que maximiza π . Note que, enquanto o nível *ótimo* de π é, por definição, seu nível *máximo*, o nível ótimo da variável de escolha Q , em si, não precisa ser nem um máximo nem um mínimo.

Para encaixar o problema em um molde mais geral e prosseguir a discussão (embora ainda nos limitando a funções objetivo de uma só variável), vamos considerar a função geral

$$y = f(x)$$

e tentar desenvolver um procedimento para achar o nível de x que maximize ou minimize o valor de y . Em nossa discussão, admitiremos que a função f é continuamente diferenciável.

9.2 Mínimo relativo e máximo relativo: teste da derivada primeira

Visto que a função objetivo $y = f(x)$ é enunciada na forma geral, não há restrições quanto a ela ser linear ou não-linear ou quanto a ser monotônica ou conter partes crescentes e decrescentes. Entre os muitos tipos possíveis de funções compatíveis com a forma da função objetivo discutida na Seção 9.1, selecionamos três casos específicos que são retratados na Figura 9.1. Embora talvez pareçam simples, os gráficos na Figura 9.1 devem nos dar uma percepção valiosa do problema de localizar o valor máximo ou mínimo da função $y = f(x)$.

Extremo relativo versus extremo absoluto

Se a função objetivo for uma função constante, como na Figura 9.1 *a*, todos os valores da variável de escolha x resultarão no mesmo valor de y , e a altura de cada ponto no gráfico da função (tal como A ou B ou C) pode ser considerada um máximo ou, também, um mínimo – ou, na verdade, nenhum dos dois. Nesse caso, não há, com efeito, nenhuma escolha significativa a fazer em relação ao valor de x para a maximização ou minimização de y .

[†] Elas também podem ser denominadas *variáveis de decisão* ou *variáveis de política*.

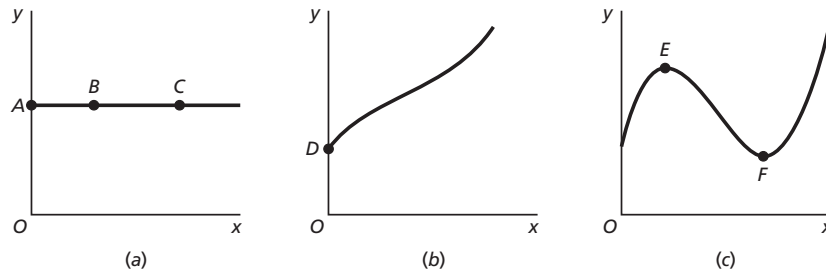


FIGURA 9.1

Na Figura 9.1b, a função é estritamente crescente e não há nenhum máximo finito se o conjunto de números reais não negativos for considerado como o domínio da função. Todavia, podemos considerar que o ponto extremo D à esquerda (a interseção com o eixo dos y) representa um mínimo; na verdade, nesse caso, ele é o mínimo *absoluto* (ou *global*) na faixa de abrangência da função.

Os pontos E e F na Figura 9.1c, por outro lado, são exemplos de um extremo *relativo* (ou *local*), no sentido de que cada um desses pontos representa um extremo apenas na vizinhança imediata do ponto. É claro que o fato de o ponto F ser um mínimo relativo não é nenhuma garantia de que ele também é o mínimo global da função, embora isso possa acontecer. De modo semelhante, um ponto máximo relativo como E pode ser ou não um máximo global. Note, também, que uma função pode muito bem ter vários extremos relativos, alguns dos quais podem ser máximos, enquanto outros são mínimos.

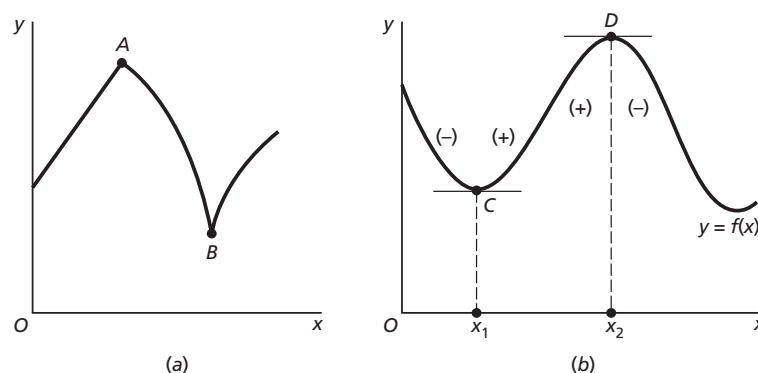
Em grande parte dos problemas econômicos que examinaremos, nossa preocupação primária, se não exclusiva, serão valores extremos que não sejam obtidos em extremidades do domínio pois, na maioria desses problemas, o domínio da função objetivo está restrito ao conjunto de números reais não-negativos e, assim, um ponto extremo (à esquerda) representará o nível zero da variável de escolha, que muitas vezes não tem nenhum interesse na prática. Na verdade, o tipo de função mais frequentemente encontrado em análise econômica é o mostrado na Figura 9.1c, ou alguma variante dele, que contém somente uma curvatura. Por conseguinte, continuaremos nossa discussão nos referindo principalmente à busca de extremos *relativos* como os pontos E e F . Isso não impedirá, de modo algum, que conheçamos um máximo absoluto se o quisermos, porque um máximo absoluto deve ser ou um máximo relativo ou um dos pontos extremos da função. Assim, se conhecermos todos os máximos relativos, bastará apenas selecionar o maior deles e compará-lo com os pontos extremos para determinar o máximo absoluto. O mínimo absoluto de uma função pode ser achado de maneira análoga. Deste ponto em diante, os valores extremos considerados serão os valores *relativos* ou *locais*, a menos que se indique o contrário.

Teste da derivada primeira

Por questão de terminologia, daqui em diante vamos nos referir à derivada de uma função alternativamente como sua derivada *primeira* (abreviatura de derivada de *primeira ordem*). A razão para isso logo ficará aparente.

Dada uma função $y = f(x)$, a derivada primeira $f'(x)$ desempenha um papel importante em nossa busca por valores extremos. Isso porque, se um extremo relativo da função ocorrer em $x = x_0$, então, (1) $f'(x_0)$ não existe ou (2) $f'(x_0) = 0$. A primeira eventualidade é ilustrada na Figura 9.2a, onde ambos os pontos, A e B , representam valores extremos relativos de y e, ainda assim, nenhuma derivada é definida em nenhum desses pontos. Posto que, na presente discussão, supomos que $y = f(x)$ é contínua e possui uma derivada contínua, contudo, na verdade, estamos excluindo pontos onde a curva forma um “bico”. No caso de funções suaves, valores extremos relativos somente podem ocorrer onde a derivada primeira tiver valor zero. Isso é ilustrado pelos pontos C e D na Figura 9.2b, ambos os quais representam valores extremos e ambos os quais são caracterizados por uma inclinação zero — $f'(x_1) = 0$ e $f'(x_2) = 0$. Também é fácil ver que, quando a inclinação for diferente de zero, não há possibilidade de termos um mínimo relativo (o fundo do vale) ou um máximo relativo (um pico). Por essa razão, no contexto de funções suaves, podemos considerar a condição $f'(x) = 0$ como uma condição *necessária* para um extremo relativo (seja máximo ou mínimo).

FIGURA 9.2



Entretanto, é preciso acrescentar que uma inclinação zero, embora *necessária*, não é *suficiente* para determinar um extremo relativo. Um exemplo de caso em que a inclinação zero não está associada a um extremo será apresentado em breve. Acrescentando uma certa cláusula à condição de inclinação zero, entretanto, podemos obter um teste decisivo para um extremo relativo, que pode ser enunciado da maneira seguinte:

Teste da derivada primeira para extremo relativo Se a derivada primeira de uma função $f(x)$ em $x = x_0$ for $f'(x_0) = 0$, então o valor da função em x_0 , $f(x_0)$, será

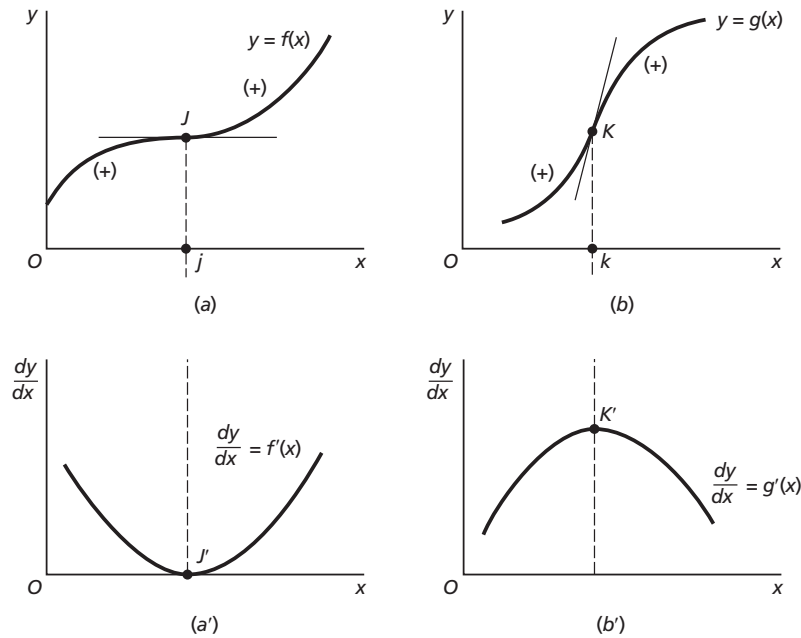
- Um *máximo* relativo se o sinal da derivada $f'(x)$ mudar de positivo para negativo da esquerda imediata do ponto x_0 até sua direita imediata.
- Um *mínimo* relativo se o sinal de $f'(x)$ mudar de negativo para positivo da esquerda imediata de x_0 até sua direita imediata.
- Nem um máximo relativo nem um mínimo relativo se $f'(x)$ tiver o mesmo sinal tanto na esquerda imediata quanto na direita imediata do ponto x_0 .

Vamos denominar o valor x_0 um *valor crítico* de x se $f'(x_0) = 0$, e denominar $f(x_0)$ como um *valor estacionário* de y (ou da função f). De acordo com isso, o ponto cujas coordenadas são x_0 e $f(x_0)$ pode ser denominado um *ponto estacionário*. (O princípio racional para a palavra *estacionário* deve ser evidente por si só – sempre que a inclinação for zero, o ponto em questão nunca estará situado sobre uma inclinação ascendente ou descendente, mas, ao contrário, estará em uma posição imutável). Então, graficamente, a primeira possibilidade relacionada nesse teste será estabelecer o ponto estacionário como o pico de uma colina, tal como o ponto D na Figura 9.2b, enquanto a segunda possibilidade estabelecerá o ponto estacionário no fundo de um vale, como o ponto C na mesma figura. Entretanto, observe que, em vista da existência de uma terceira possibilidade, ainda não discutida, não podemos considerar a condição $f'(x) = 0$ como uma *condição suficiente* para um máximo ou um mínimo relativo. Porém, agora vemos que, se a condição necessária $f'(x) = 0$ for satisfeita, então a cláusula de mudança de sinal da derivada pode servir como uma *condição suficiente* para um máximo ou um mínimo relativo, dependendo da direção da mudança de sinal.

Agora vamos explicar a terceira possibilidade. A Figura 9.3a mostra que a função f atinge uma inclinação zero no ponto J (quando $x = j$). Embora $f'(j)$ seja zero – o que torna $f(j)$ um valor estacionário – a derivada não muda de sinal de um lado para o outro de $x = j$; por conseguinte, de acordo com o teste da derivada primeira, o ponto J não é nem máximo nem mínimo, o que é devidamente confirmado pelo gráfico da função. Em vez disso, ele é um exemplo do que é conhecido como um *ponto de inflexão*.

A característica fundamental de um ponto de inflexão é que, nesse ponto, a função derivada (no sentido oposto de função primitiva) alcança um valor extremo. Posto que esse valor extremo pode ser um máximo ou um mínimo, temos dois tipos de pontos de inflexão. Na Figura 9.3a', onde representamos graficamente a derivada $f'(x)$, vemos que esse valor é zero quando $x = j$ (veja o ponto J'), mas é positivo de ambos os lados do ponto J' ; isso faz de J' um ponto de *mínimo* da função derivada $f'(x)$.

FIGURA 9.3



O outro tipo de ponto de inflexão é retratado na Figura 9.3b, onde a inclinação da função $g(x)$ aumenta até alcançar o ponto k e, então, passa a decrescer. Por consequência, o gráfico da função derivada $g'(x)$ assumirá a forma mostrada na Figura 9.3b', onde o ponto K' dá um valor *máximo* da função derivada $g'(x)$.[†]

Resumindo: um extremo relativo deve ser um valor estacionário, mas um valor estacionário pode estar associado a um extremo relativo ou a um ponto de inflexão. Para achar o máximo ou o mínimo relativo de uma dada função, portanto, o procedimento deve ser, em primeiro lugar, achar os valores estacionários da função onde a condição $f'(x) = 0$ é satisfeita e, então, aplicar o teste da derivada primeira para determinar se cada um dos valores estacionários é um máximo relativo, um mínimo relativo ou nenhum dos dois.

EXEMPLO 1 Encontre os extremos relativos da função

$$y = f(x) = x^3 - 12x^2 + 36x + 8$$

Em primeiro lugar, encontramos a função derivada, ou seja

$$f'(x) = 3x^2 - 24x + 36$$

Para obter os valores críticos, isto é, os valores de x que satisfazem a condição $f'(x) = 0$, igualamos a função derivada quadrática a zero e obtemos a equação quadrática

$$3x^2 - 24x + 36 = 0$$

Fatorando o polinômio ou aplicando a fórmula quadrática, obtemos, então, o seguinte par de raízes (soluções):

$$x_1^* = 6 \quad [\text{no qual temos } f'(6) = 0 \text{ e } f(6) = 8]$$

$$x_2^* = 2 \quad [\text{no qual temos } f'(2) = 0 \text{ e } f(2) = 40]$$

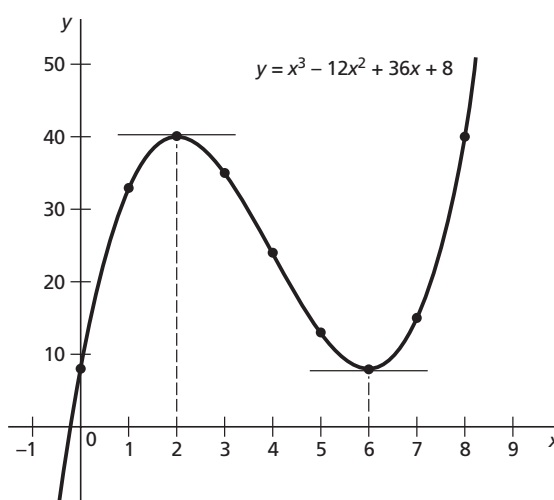
Visto que $f'(6) = f'(2) = 0$, esses dois valores de x são os valores críticos que desejamos.

[†] Note que um valor zero da derivada, embora seja uma condição necessária para um extremo relativo, não é exigido para um ponto de inflexão; pois a derivada $g'(x)$ tem um valor positivo em $x = k$ e, ainda assim, o ponto K é um ponto de inflexão.

É fácil verificar que, na vizinhança imediata de $x = 6$, temos $f'(x) < 0$ para $x < 6$, e $f'(x) > 0$ para $x > 6$; assim, o valor da função $f(6) = 8$ é um mínimo relativo. De modo semelhante, visto que na vizinhança imediata de $x = 2$ achamos $f'(x) > 0$ para $x < 2$, e $f'(x) < 0$ para $x > 2$, o valor da função $f(2) = 40$ é um máximo relativo.

A Figura 9.4 mostra o gráfico da função deste exemplo. Esse gráfico pode ser usado para verificar a localização dos valores extremos obtidos por meio da utilização do teste da derivada primeira. Mas, na realidade, na maioria dos casos, acontece o contrário – os valores extremos derivados matematicamente ajudarão na construção do gráfico. Para construir um gráfico com precisão, o ideal seria conhecer o valor da função em todos os pontos do domínio; mas, na prática, apenas alguns pontos do domínio são selecionados para a finalidade de traçar o gráfico e os pontos restantes normalmente são determinados por interpolação. O perigo dessa prática é que, a menos que encontremos o ponto (ou os pontos) estacionário(s) por coincidência, não acharemos a exata localização do ponto (ou pontos) de reversão da curva. Agora, tendo o teste da derivada primeira à nossa disposição, é possível localizar com precisão esses pontos de reversão.

FIGURA 9.4

**EXEMPLO 2**

Encontre o extremo relativo da função de custo médio

$$AC = f(Q) = Q^2 - 5Q + 8$$

Aqui, a derivada é $f'(Q) = 2Q - 5$, uma função linear. Igualando $f'(Q)$ a zero, obtemos a equação linear $2Q - 5 = 0$, que tem uma só raiz, $Q^* = 2,5$, que é o único valor crítico nesse caso. Para aplicar o teste da derivada primeira, vamos encontrar os valores da derivada em, digamos, $Q = 2,4$ e $Q = 2,6$, respectivamente. Visto que $f'(2,4) = -0,2 < 0$, enquanto $f'(2,6) = 0,2 > 0$, podemos concluir que o valor estacionário $AC = f(2,5) = 1,75$ representa um mínimo relativo. O gráfico da função desse exemplo é, na verdade, uma curva em U, de modo que o mínimo relativo já encontrado também será o mínimo absoluto. Conhecer a localização exata desse ponto deverá ser de grande auxílio para traçar a curva AC.

EXERCÍCIO 9.2

1. Encontre os valores estacionários das seguintes funções (verifique se são máximos ou mínimos relativos ou pontos de inflexão), admitindo que o domínio seja o conjunto de todos os números reais:

(a) $y = -2x^2 + 8x + 7$ (b) $y = 5x^2 + x$ (c) $y = 3x^2 + 3$ (d) $y = 3x^2 - 6x + 2$

2. Encontre os valores estacionários das seguintes funções (verifique se são máximos ou mínimos relativos ou pontos de inflexão), admitindo que o domínio seja o intervalo $[0, \infty)$:
 - (a) $y = x^3 - 3x + 5$
 - (b) $y = \frac{1}{3}x^3 - x^2 + x + 10$
 - (c) $y = -x^3 + 4,5x^2 - 6x + 6$
3. Mostre que a função $y = x + 1/x$ (com $x \neq 0$) tem dois extremos relativos, sendo um deles um máximo e o outro um mínimo. O "mínimo" é maior ou menor que o "máximo"? Como este resultado paradoxal é possível?
4. Seja $T = \phi(x)$ uma função *total* (por exemplo, produção total ou custo total):
 - (a) Escreva as expressões para a função *marginal* M e para a função *média* A .
 - (b) Mostre que, quando A alcança um extremo relativo, M e A devem ter o mesmo valor.
 - (c) Isso sugere qual princípio geral para o desenho de uma curva marginal e de uma curva média no mesmo diagrama?
 - (d) Qual sua conclusão sobre a elasticidade da função total T no ponto onde A alcança um valor extremo?

9.3 Derivadas segundas e derivadas de ordens mais altas

Até aqui, consideramos apenas a derivada primeira $f'(x)$ de uma função $y = f(x)$; agora, vamos introduzir o conceito de *derivada segunda* (abreviação de *derivada de segunda ordem*) e derivadas de ordens ainda mais altas. Isso nos habilitará a desenvolver critérios alternativos para a localização dos extremos relativos de uma função.

Derivada de uma derivada

Visto que a derivada primeira $f'(x)$ é, em si, uma função de x , ela também deve ser diferenciável em relação a x , contanto que seja contínua e suave. O resultado dessa diferenciação, conhecido como a derivada segunda da função f , é denotado por

$f''(x)$ onde as aspas duplas (duas linhas) indicam que $f(x)$ foi diferenciada duas vezes em relação a x , e onde a expressão (x) após as duas linhas (aspas duplas) sugere que a derivada segunda é, mais uma vez, uma função de x

ou

$\frac{d^2 y}{dx^2}$ onde a notação decorre da consideração de que a derivada segunda significa, de fato, $\frac{d}{dx} \left(\frac{dy}{dx} \right)$; daí o d^2 (lê-se "d dois") no numerador e o dx^2 (lê-se "dx ao quadrado") no denominador deste símbolo.

Se a derivada segunda $f''(x)$ existir para todos os valores de x no domínio, diz-se que a função $f(x)$ é *duplamente diferenciável*; se, além disso, $f''(x)$ for contínua, diz-se que a função $f(x)$ é *duplamente diferenciável continuamente*. Exatamente como a notação $f \in C^{(1)}$ ou $f \in C'$ é freqüentemente usada para indicar que a função f é continuamente diferenciável, uma notação análoga

$$f \in C^{(2)} \quad \text{ou} \quad f \in C''$$

pode ser usada para significar que f é duplamente diferenciável continuamente.

Por ser uma função de x , a derivada segunda pode ser diferenciada em relação a x mais uma vez, produzindo a derivada *terceira*, que, por sua vez, pode ser a fonte de uma derivada *quarta* e assim por diante, *ad infinitum*, desde que satisfeitas as condições de diferenciabilidade. Essas derivadas de ordens mais altas são simbolizadas segundo as mesmas linhas da derivada segunda:

$$f'''(x), f^{(4)}(x), \dots, f^{(n)}(x) \quad [\text{com índices entre } ()]$$

ou

$$\frac{d^3 y}{dx^3}, \frac{d^4 y}{dx^4}, \dots, \frac{d^n y}{dx^n}$$

A última delas também pode ser escrita como $\frac{d^n}{dx^n} y$, onde a parte $\frac{d^n}{dx^n}$ serve como um símbolo de operador que nos indica que devemos tomar a n -ésima derivada de (alguma função) em relação a x .

Quase todas as funções *específicas* com as quais trabalharemos possuem derivadas contínuas até qualquer ordem que desejarmos; isto é, elas são continuamente diferenciáveis qualquer número de vezes. Toda vez que for usada uma função *geral*, tal como $f(x)$, sempre vamos supor que ela tem derivadas até qualquer ordem que precisarmos.

EXEMPLO 1 Encontre as derivadas da primeira até a quinta da função

$$y = f(x) = 4x^4 - x^3 + 17x^2 + 3x - 1$$

As derivadas desejadas são as seguintes:

$$f'(x) = 16x^3 - 3x^2 + 34x + 3$$

$$f''(x) = 48x^2 - 6x + 34$$

$$f'''(x) = 96x - 6$$

$$f^{(4)}(x) = 96$$

$$f^{(5)}(x) = 0$$

Nesse exemplo específico (polinomial), notamos que cada função derivada sucessiva surge como um polinômio de ordem mais baixa – de cúbico a quadrático, a linear, a constante. Notamos, também, que a derivada quinta, sendo a derivada de uma constante, é igual a zero para todos os valores de x ; portanto, também poderíamos tê-la escrito como $f^{(5)}(x) \equiv 0$. A equação $f^{(5)}(x) = 0$ deve ser cuidadosamente distinguida da equação $f^{(5)}(x_0) = 0$ (zero em x_0 somente). Além disso, entenda que a declaração $f^{(5)}(x) \equiv 0$ não significa que a derivada quinta não existe; na verdade, ela existe e seu valor é zero.

EXEMPLO 2 Encontre as quatro primeiras derivadas da função racional

$$y = g(x) = \frac{x}{1+x} \quad (x \neq -1)$$

Essas derivadas podem ser encontradas por meio da utilização da regra do quociente ou, após reescrever a função como $y = x(1+x)^{-1}$, pela regra do produto:

$$\left. \begin{aligned} g'(x) &= (1+x)^{-2} \\ g''(x) &= -2(1+x)^{-3} \\ g'''(x) &= 6(1+x)^{-4} \\ g^{(4)}(x) &= -24(1+x)^{-5} \end{aligned} \right\} \quad (x \neq -1)$$

Nesse caso, é evidente que a derivação repetida não tende a simplificar as expressões derivadas subsequentes.

Note que, do mesmo modo que a função primitiva $g(x)$, todas as derivadas sucessivas obtidas são, em si mesmas, funções de x . Entretanto, dados valores específicos de x , essas funções derivadas assumirão valores específicos. Quando $x = 2$, por exemplo, a derivada segunda no Exemplo 2 pode ser calculada como

$$g''(2) = -2(3)^{-3} = \frac{-2}{27}$$

e de modo semelhante para outros valores de x . É de suprema importância entender que, para avaliar essa derivada segunda $g''(x)$ em $x = 2$ como fizemos, devemos, em primeiro lugar, obter $g''(x)$ de $g'(x)$ e então substituir $x = 2$ na equação para $g''(x)$. É *incorreto* substituir $x = 2$ em $g(x)$ ou $g'(x)$ antes do processo de diferenciação que leva a $g''(x)$.

Interpretação da derivada segunda

A função derivada $f'(x)$ mede a taxa de variação da função f . Pelo mesmo critério, a função derivada segunda f'' é a medida da taxa de variação da derivada primeira f' ; em outras palavras, a derivada segunda mede a *taxa de variação da taxa de variação* da função original f . Ou, com outro enunciado, dado um aumento infinitesimal na variável independente x a partir de um ponto $x = x_0$,

$$\left. \begin{array}{l} f'(x_0) > 0 \\ f'(x_0) < 0 \end{array} \right\} \text{significa que o valor da função tende a } \begin{cases} \text{aumentar} \\ \text{decrecer} \end{cases}$$

ao passo que, no que se refere à derivada segunda,

$$\left. \begin{array}{l} f''(x_0) > 0 \\ f''(x_0) < 0 \end{array} \right\} \text{significa que a inclinação da curva tende a } \begin{cases} \text{aumentar} \\ \text{decrecer} \end{cases}$$

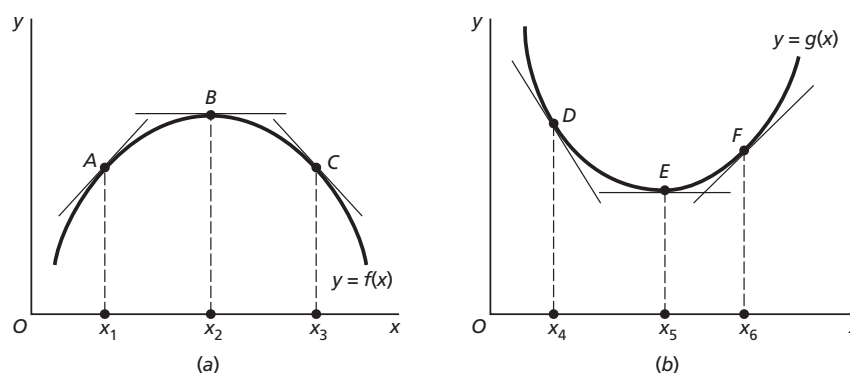
Assim, uma derivada primeira positiva conjugada com uma derivada segunda positiva em $x = x_0$ implica que a inclinação da curva naquele ponto é *positiva e crescente*. Em outras palavras, o valor da função está aumentando a uma taxa crescente. Do mesmo modo, uma derivada primeira positiva conjugada com uma derivada segunda negativa indica que a inclinação da curva é *positiva mas decrescente* – o valor da função está aumentando a uma taxa decrescente. O caso de uma derivada primeira negativa pode ser interpretado de modo análogo, mas, nesse caso, uma advertência é oportuna: quando $f'(x_0) < 0$ e $f''(x_0) > 0$, a inclinação da curva é *negativa e crescente*, mas isso *não* significa que a inclinação está mudando, digamos, de (-10) para (-11) ; ao contrário, a variação deve ser de (-11) , um número menor, para (-10) , um número maior. Em outras palavras, a inclinação da curva deve tender a ser *menos íngreme* à medida que x aumenta. Por fim, quando $f'(x_0) < 0$ e $f''(x_0) < 0$, a inclinação da curva deve ser *negativa e decrescente*. Isso se refere a uma inclinação negativa que tende a ficar *mais íngreme* à medida que x aumenta.

Tudo isso pode ficar mais claro com uma explicação gráfica. A Figura 9.5a ilustra uma função com $f''(x) < 0$ em toda sua extensão. Uma vez que a inclinação deve decrescer regularmente à medida que x aumenta no gráfico, quando atravessarmos da esquerda para a direita, passaremos por um ponto A cuja inclinação é positiva, em seguida por um ponto B cuja inclinação é zero e, então, por um ponto C cuja inclinação é negativa. É claro que pode acontecer de uma função com $f''(x) < 0$ ser caracterizada por $f'(x) > 0$ em toda sua extensão e, por isso, aparecer no gráfico apenas como a porção crescente de uma curva em forma de U, ou por $f'(x) < 0$ em toda a sua extensão e então aparecer no gráfico apenas como a porção decrescente daquela curva.

O caso oposto de uma função com $f''(x) > 0$ em toda a sua extensão é ilustrado na Figura 9.5b. Aqui, ao passarmos pelos pontos D a E a F , a inclinação cresce constantemente e muda de negativa para zero para positiva. Mais uma vez, advertimos que, dependendo da especificação da derivada primeira, uma função caracterizada por $f''(x) > 0$ em toda a sua extensão pode aparecer no gráfico como as porções crescentes ou decrescentes de uma curva em U.

Fica evidente, pela Figura 9.5, que a derivada segunda $f''(x)$ está relacionada à *curvatura* de um gráfico; ela determina como a curva tende a se curvar. Para descrever os dois tipos diferentes de curvatura discutidos, vamos nos referir à curva na Figura 9.5a como *estritamente côncava* e à da Figura 9.5b como *estritamente convexa*. E, bem entendido, uma função cujo gráfico é estritamente côncavo (estritamente convexo) é denominada uma *função estritamente côncava* (*estritamente convexa*). A caracterização geométrica precisa de uma função estritamente côncava é a seguinte: se escolhermos qualquer par de pontos M e N sobre seu gráfico e os unirmos por uma linha reta, o segmento de reta MN deve estar inteiramente *abaixo* da curva, exceto nos pontos M e N . A carac-

FIGURA 9.5



terização de uma função estritamente convexa pode ser obtida substituindo-se a palavra *acima* por *abaixo* na última sentença. Experimente fazer isso na Figura 9.5. Se abrandarmos um pouco a condição de caracterização de modo a permitir que o segmento de reta MN esteja *ou* abaixo da curva *ou* ao longo da curva (coincidindo com ela), então estaremos descrevendo uma *função côncava*, sem o advérbio *estritamente*. De maneira semelhante, se o segmento de reta MN ficar *ou* acima *ou* ao longo da curva, então a função é *convexa*, novamente sem o advérbio *estritamente*. Note que, visto que o segmento de reta MN pode coincidir com uma curva (não estritamente) côncava ou convexa, a última pode muito bem conter um segmento linear. Ao contrário, uma curva *estritamente* côncava ou convexa nunca pode conter um segmento linear em lugar algum. Deduz-se que, enquanto uma função estritamente côncava (convexa) é automaticamente uma função côncava (convexa), o inverso não é verdade.[†]

Com base em nossa discussão anterior da derivada segunda, agora podemos inferir que, se a derivada segunda $f''(x)$ for negativa para *todo* x , então a função primitiva $f(x)$ deve ser uma função estritamente côncava. De modo semelhante, $f(x)$ deve ser estritamente convexa se $f''(x)$ for positiva para *todo* x . Apesar disso, *não* é válido inverter essa inferência e dizer que, se $f(x)$ for estritamente côncava (estritamente convexa), então $f''(x)$ deve ser negativa (positiva) para todo x . Isso porque, em certos casos excepcionais, a derivada segunda pode ter um valor zero em um ponto estacionário de tal curva. Um exemplo disso pode ser encontrado na função $y = f(x) = x^4$, cujo gráfico é uma curva estritamente convexa, mas cujas derivadas

$$f'(x) = 4x^3 \quad f''(x) = 12x^2$$

indicam que, no ponto estacionário onde $x = 0$, o valor da derivada segunda é $f''(0) = 0$. Contudo, observe que, em qualquer outro ponto, com $x \neq 0$, a derivada segunda dessa função tem, na verdade, o sinal positivo (esperado). Por conseguinte, à parte a possibilidade de um valor zero em um ponto estacionário, pode-se esperar que, em geral, a derivada segunda de uma função estritamente côncava ou convexa conserve um único sinal algébrico.

Para outros tipos de função, a derivada segunda pode assumir valores positivos e negativos, dependendo do valor de x . Nas Figuras 9.3a e b, por exemplo, ambas, $f(x)$ e $g(x)$ sofrem uma mudança de sinal na derivada segunda em seus respectivos pontos de inflexão J e K . De acordo com a Figura 9.3a', a inclinação de $f'(x)$ – isto é, o valor de $f''(x)$ – muda de negativo para positivo em $x = j$; ocorre exatamente o oposto na inclinação de $g'(x)$ – isto é, o com valor de $g''(x)$ – segundo a Figura 9.3b'. Traduzindo em termos de curvatura, isto significa que o gráfico de $f(x)$ passa de estritamente côncavo para estritamente convexo no ponto J , ao passo que ocorre a mudança inversa no gráfico de $g(x)$ no ponto K . Por conseqüência, em vez de caracterizar um ponto de inflexão como um ponto em que a derivada primeira alcança um valor extremo, podemos caracterizá-lo alternativamente como um ponto onde a função sofre uma mudança na sua curvatura ou uma mudança no sinal de sua derivada segunda.

[†] Trataremos desses conceitos mais a fundo na Seção 11.5.

Uma aplicação

As duas curvas na Figura 9.5 exemplificam os gráficos de funções quadráticas, as quais podem ser expressas de modo geral na forma

$$y = ax^2 + bx + c \quad (a \neq 0)$$

Tendo como base nossa discussão da derivada segunda, agora podemos derivar um modo conveniente de determinar se uma dada função quadrática terá um gráfico estritamente convexo (em forma de U) ou estritamente côncavo (em forma de U invertido).

Visto que a derivada segunda da função quadrática citada é $d^2y/dx^2 = 2a$, essa derivada sempre terá o mesmo sinal algébrico que a coeficiente a . Recordando que uma derivada segunda positiva implica uma curva estritamente convexa, podemos inferir que um coeficiente a positivo na função quadrática precedente dá origem a um gráfico em forma de U. Ao contrário, um coeficiente a negativo resulta em uma curva estritamente côncava, na forma de um U invertido.

Como demos a entender no final da Seção 9.2, verificaremos que o extremo relativo dessa função também será seu extremo absoluto porque em uma função quadrática, somente pode ser encontrado um único vale ou pico, que fica evidente em um U ou em um U invertido, respectivamente.

Atitudes em relação ao risco

A aplicação mais comum do conceito de utilidade marginal é no contexto do consumo de bens. Mas, em uma outra aplicação útil, consideramos a utilidade marginal da *renda* ou, mais especificamente em relação à presente discussão, o *retorno* de um jogo de azar, e usamos esse conceito para distinguir entre as diferentes atitudes de indivíduos em relação ao risco.

Considere o jogo no qual, mediante uma quantia fixa de dinheiro paga adiantadamente (o *custo* do jogo), você pode lançar um dado e ganhar \$10 se sair um número ímpar ou \$20 se sair um número par. Em vista da igual probabilidade dos dois resultados, o *valor esperado do retorno* é, matematicamente

$$EV = 0,5 \times \$10 + 0,5 \times \$20 = \$15$$

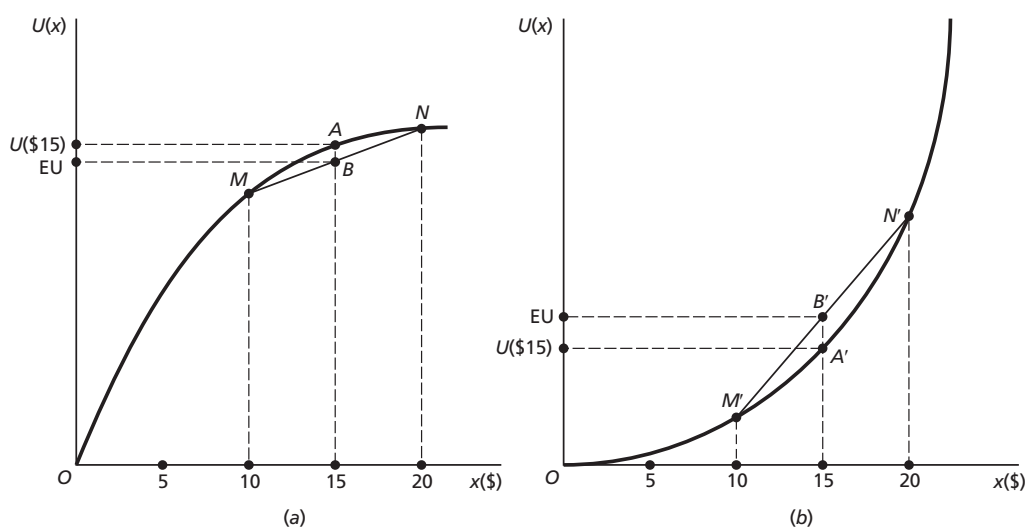
O jogo é considerado um *jogo justo*, ou uma *aposta justa*, se seu custo for exatamente \$15. A despeito de sua justiça, participar de um jogo como esse envolve um risco, pois, mesmo que a distribuição de probabilidade dos dois resultados possíveis seja conhecida, o resultado propriamente dito de qualquer jogada individual não é. Daí, pessoas “adversas ao risco” sempre se negariam a jogá-lo. Por outro lado, há pessoas que “gostam do risco” ou que “preferem o risco” e que adorariam participar de jogos justos ou até mesmo de jogos cujas probabilidades lhes são adversas (isto é, quando o custo do jogo é maior que o valor do retorno).

A explicação para essas atitudes diversas em relação ao risco é encontrada com facilidade nas diferentes funções utilidade que as pessoas possuem. Suponha que um jogador potencial tenha a função utilidade estritamente côncava $U = U(x)$ representada na Figura 9.6a, onde x denota o retorno, com $U(0) = 0$, $U'(x) > 0$ (utilidade marginal positiva da renda ou do retorno) e $U''(x) < 0$ (utilidade marginal decrescente) para todo x . A decisão econômica que essa pessoa enfrenta envolve a escolha entre dois cursos de ação: primeiro, se não jogar, ela economiza os \$15 do custo do jogo (= EV) e, assim, desfruta o nível de utilidade $U(\$15)$, medido pela altura do ponto A na curva. Segundo, se jogar, ela tem uma probabilidade de 0,5 de receber \$10 e, assim, desfrutar $U(\$10)$ (veja ponto M), mais uma probabilidade de 0,5 de receber \$20 e, assim, desfrutar $U(\$20)$ (veja ponto N). A *utilidade esperada de jogar* é, portanto, igual a

$$EU = 0,5 \times U(\$10) + 0,5 \times U(\$20)$$

a qual, por ser a média entre as alturas de M e de N, é medida pela altura do ponto B, o ponto médio do segmento de reta MN. Uma vez que, pela propriedade que define uma função utilidade estritamente côncava, o segmento de reta MN tem de estar abaixo do arco MN, o ponto B deve ser mais baixo que o ponto A; isto é, EU, a utilidade esperada de jogar fica abaixo do custo de uti-

FIGURA 9.6



lidade do jogo e ele seria evitado. Por essa razão, uma função utilidade estritamente côncava é associada a um comportamento de aversão ao risco.

Para uma pessoa que gosta de arriscar, o processo de decisão é análogo, mas a decisão tomada será a oposta porque, agora, a função utilidade relevante é estritamente convexa. Na Figura 9.6b, $U(\$15)$, a utilidade de conservar os \$15 por não jogar, é mostrada pelo ponto A' na curva, e EU , a utilidade esperada de jogar, é dada por B' , o ponto médio sobre o segmento de reta $M'N'$. Mas, dessa vez, o segmento de reta $M'N'$ está acima do arco $M'N'$, e o ponto B' está acima do ponto A' . Assim, há, definitivamente, um incentivo positivo para jogar. Ao contrário da situação na Figura 9.6a, desse modo podemos associar uma função utilidade estritamente convexa a um comportamento favorável ao risco.

EXERCÍCIO 9.3

1. Encontre a segunda e a terceira derivadas das seguintes funções:

(a) $ax^2 + bx + c$ (c) $\frac{3x}{1-x}$ ($x \neq 1$)

(b) $7x^4 - 3x - 4$ (d) $\frac{1+x}{1-x}$ ($x \neq 1$)

2. Quais das seguintes funções quadráticas são estritamente convexas?

(a) $y = 9x^2 - 4x + 8$ (c) $u = 9 - 2x^2$
(b) $w = -3x^2 + 39$ (d) $v = 8 - 5x + x^2$

3. Desenhe (a) uma curva côncava que *não* é estritamente côncava e (b) uma curva que se qualifica simultaneamente como uma curva côncava e uma curva convexa.

4. Dada a função $y = a - \frac{b}{c+x}$ ($a, b, c > 0$; $x \geq 0$), determine a forma geral de seu gráfico examinando (a) sua primeira e segunda derivadas; (b) sua interseção com o eixo vertical; e (c) o limite de y quando x tende ao infinito. Se essa função fosse usada como uma função de consumo, que restrições deveriam ser aplicadas aos parâmetros para torná-la economicamente razoável?

5. Desenhe o gráfico de uma função $f(x)$ tal que $f'(x) \equiv 0$, e o gráfico de uma função $g(x)$ tal que $g'(3) = 0$. Resuma em uma sentença a diferença essencial entre $f(x)$ e $g(x)$ em termos do conceito de ponto estacionário.

6. Diz-se que uma pessoa que não é nem favorável nem adversa ao risco (indiferente a um jogo justo) é "neutra em relação ao risco".

(a) Que tipo de função utilidade você usaria para caracterizar tal pessoa?

(b) Usando o jogo de dados detalhado no texto, descreva a relação entre $U(\$15)$ e EU para a pessoa neutra em relação ao risco.

9.4 Teste da derivada segunda

Voltando ao par de pontos extremos B e E na Figura 9.5 e lembrando a relação recém-estabelecida entre a derivada segunda e a curvatura de uma curva, podemos constatar a validade dos seguintes critérios para um extremo relativo:

Teste da derivada segunda para extremo relativo Se o valor da derivada primeira de uma função f em $x = x_0$ for $f'(x_0) = 0$, então o valor da função em x_0 , $f(x_0)$, será

- Um *máximo* relativo se o valor da derivada segunda em x_0 for $f''(x_0) < 0$.
- Um *mínimo* relativo se o valor da derivada segunda em x_0 for $f''(x_0) > 0$.

Esse teste é, em geral, mais conveniente de usar do que o teste da derivada primeira, porque não requer que verifiquemos o sinal da derivada à esquerda e também à direita de x_0 . Porém, tem a desvantagem de não se poder tirar nenhuma conclusão inequívoca quando $f''(x_0) = 0$. Então, o valor estacionário $f(x_0)$ pode ser *ou* um máximo relativo *ou* um mínimo relativo *ou* até mesmo um valor de inflexão.[†] Quando encontramos a situação de $f''(x_0) = 0$, devemos voltar ao teste da derivada primeira ou recorrer a um outro teste, que será desenvolvido na Seção 9.6, e que envolve a derivada terceira ou até derivadas de ordens mais altas. Para grande parte dos problemas econômicos, entretanto, o teste da derivada segunda usualmente seria adequado para determinar um máximo relativo ou um mínimo relativo.

EXEMPLO 1 Encontre o extremo relativo da função

$$y = f(x) = 4x^2 - x$$

As derivadas primeira e segunda são

$$f'(x) = 8x - 1 \quad \text{e} \quad f''(x) = 8$$

Igualando $f'(x)$ a zero e resolvendo a equação resultante, verificamos que o (único) valor crítico é $x^* = \frac{1}{8}$, o que dá o (único) valor estacionário $f\left(\frac{1}{8}\right) = -\frac{1}{16}$. Como a derivada segunda é positiva (nesse caso ela é, de fato, positiva para qualquer valor de x), o extremo é estabelecido como um mínimo. Além disso, visto que o gráfico da função dada tem a forma de uma curva em U, o mínimo relativo também é o mínimo absoluto.

EXEMPLO 2 Encontre os extremos relativos da função

$$y = g(x) = x^3 - 3x^2 + 2$$

As primeiras duas derivadas dessa função são

$$g'(x) = 3x^2 - 6x \quad \text{e} \quad g''(x) = 6x - 6$$

Igualando $g'(x)$ a zero e resolvendo a equação quadrática resultante, $3x^2 - 6x = 0$, obtemos os valores críticos $x_1^* = 2$ e $x_2^* = 0$, que, por sua vez, resultam nos dois valores estacionários:

$$g(2) = -2 \quad [\text{um mínimo porque } g''(2) = 6 > 0]$$

$$g(0) = 2 \quad [\text{um máximo porque } g''(0) = -6 < 0]$$

[†] Para verificar que um ponto de inflexão é possível quando $f''(x_0) = 0$, vamos voltar às Figuras 9.3a e 9.3a'. O ponto J no diagrama superior é um ponto de inflexão, cujo valor crítico é $x = j$. Visto que o gráfico de $f'(x)$ no diagrama inferior atinge um mínimo em $x = j$, a inclinação de $f'(x)$ [isto é, $f''(x)$] tem de ser zero no valor crítico $x = j$. Assim, o ponto J ilustra um ponto de inflexão que ocorre quando $f''(x_0) = 0$.

Para verificar que um extremo relativo também é consistente com $f''(x_0) = 0$, considere a função $y = x^4$. O gráfico dessa função é uma curva em U e tem um mínimo $y = 0$, atingido no valor crítico $x = 0$. Visto que a derivada segunda dessa função é $f''(x) = 12x^2$, novamente obtemos um valor zero para essa derivada no valor crítico $x = 0$. Assim, essa função ilustra um extremo relativo que ocorre quando $f''(x_0) = 0$.

Condições necessárias versus condições suficientes

Do mesmo modo que no teste da derivada primeira, a condição de inclinação zero $f'(x) = 0$ desempenha o papel de uma condição *necessária* no teste da derivada segunda. Uma vez que essa condição é baseada na derivada de primeira ordem, ela é freqüentemente denominada *condição de primeira ordem*. Uma vez verificado que a condição de primeira ordem é satisfeita em $x = x_0$, o sinal negativo (positivo) de $f''(x_0)$ é *suficiente* para estabelecer o valor estacionário em questão como um máximo (mínimo) relativo. Essas condições suficientes, que são baseadas na derivada de segunda ordem, são geralmente denominadas *condições de segunda ordem*.

Cabe repetir que a condição de primeira ordem é *necessária*, mas não *suficiente*, para um máximo ou um mínimo relativo. (Lembra-se dos pontos de inflexão?) Contrastando acentuadamente com isso, a condição de que $f''(x)$ seja negativa (positiva) no valor crítico x_0 é *suficiente* para um máximo (mínimo) relativo, mas *não é necessária*. [Lembra-se do extremo relativo que ocorre quando $f''(x_0) = 0$?] Por essa razão, é preciso se precaver cuidadosamente contra a seguinte linha de argumentação: “Uma vez que já se sabe que o valor $f(x_0)$ é um mínimo, devemos ter $f''(x_0) > 0$ ”. Este raciocínio é falho porque considera incorretamente o sinal positivo de $f''(x_0)$ como uma condição necessária para que $f(x_0)$ seja um mínimo.

Isso não quer dizer que derivadas de segunda ordem nunca podem ser usadas para determinar condições *necessárias* para extremos relativos. Na verdade, podem. Mas é preciso não esquecer de levar em conta o fato de que um máximo (mínimo) relativo pode ocorrer não apenas quando $f''(x_0)$ é negativo (positivo), mas também quando $f''(x_0)$ é zero. Por consequência, *condições necessárias de segunda ordem* devem ser enunciadas em termos de desigualdades fracas: para que um valor estacionário $f(x_0)$ seja um $\begin{cases} \text{máximo} \\ \text{mínimo} \end{cases}$ relativo, é necessário que $f''(x_0) \begin{cases} \leq \\ \geq \end{cases} 0$.

A discussão precedente pode ser resumida na Tabela 9.1. Todas as equações e desigualdades na tabela têm as características de condições (requisitos) a serem cumpridos, em vez de especificações descritivas de uma dada função. Em particular, a equação $f'(x) = 0$ não significa que a inclinação da função f seja igual a zero em toda a sua extensão; em vez disso, ela determina que somente os valores de x que satisfazem esse requisito podem se qualificar como valores críticos.

TABELA 9.1
Condições para
um extremo
relativo: $y = f(x)$

Condição	Máximo	Mínimo
Necessária de primeira ordem	$f'(x) = 0$	$f'(x) = 0$
Necessária de segunda ordem [†]	$f''(x) \leq 0$	$f''(x) \geq 0$
Suficiente de segunda ordem [†]	$f''(x) < 0$	$f''(x) > 0$

[†]Aplicável somente após satisfazer a condição necessária de primeira ordem.

Condições para maximização de lucro

Apresentaremos agora um exemplo econômico de problemas de valores extremos, isto é, problemas de otimização.

Uma das primeiras coisas que um estudante de economia aprende é que, para maximizar o lucro, uma empresa deve igualar custo marginal e receita marginal. Vamos mostrar a derivação matemática dessa condição. Para manter a análise em um nível geral, vamos trabalhar com a função de receita total $R = R(Q)$ e a função de custo total $C = C(Q)$, ambas funções de uma única variável Q . Delas deduzimos que uma função de lucro (a função objetivo) também pode ser formulada em termos de Q (a variável de escolha):

$$\pi = \pi(Q) = R(Q) - C(Q) \quad (9.1)$$

Para achar o nível de produção que maximiza o lucro, devemos satisfazer a condição necessária de primeira ordem para um máximo: $d\pi/dQ = 0$. Portanto, vamos diferenciar (9.1) em relação a Q e igualar a derivada resultante a zero. O resultado é:

$$\begin{aligned}\frac{d\pi}{dQ} &\equiv \pi'(Q) = R'(Q) - C'(Q) \\ &= 0 \text{ se e somente se } R'(Q) = C'(Q)\end{aligned}\quad (9.2)$$

Assim, a produção *ótima* (produção *de equilíbrio*) Q^* deve satisfazer a equação $R'(Q^*) = C'(Q^*)$ ou $MR = MC$. Essa condição constitui a condição de primeira ordem para maximização de lucro.

Contudo, a condição de primeira ordem pode levar a um mínimo, em vez de um máximo; por isso, temos de verificar a condição de segunda ordem em seguida. Podemos obter a derivada segunda diferenciando a derivada primeira em (9.2) em relação a Q :

$$\begin{aligned}\frac{d^2\pi}{dQ^2} &\equiv \pi''(Q) = R''(Q) - C''(Q) \\ &\leq 0 \text{ se e somente se } R''(Q) \leq C''(Q)\end{aligned}$$

Esta última desigualdade é a condição necessária de segunda ordem para a maximização. Se não for cumprida, então Q^* não tem possibilidade de maximizar o lucro; na verdade, ela o minimiza. Se $R''(Q^*) = C''(Q^*)$, então não podemos chegar a uma conclusão definitiva. O melhor cenário é achar $R''(Q^*) < C''(Q^*)$, que satisfaz a condição suficiente de segunda ordem para um máximo. Nesse caso, podemos concluir que Q^* é uma produção que maximiza o lucro. Em termos econômicos, isso significaria que, se a taxa de variação de MR for menor que a taxa de variação de MC na produção em que $MC = MR$, então a produção maximizará o lucro.

Essas condições estão ilustradas na Figura 9.7. Na Figura 9.7a, desenhamos uma curva de receita total e uma curva de custo total, que, podemos observar, interceptam-se duas vezes – nos níveis de produção Q_2 e Q_4 . No intervalo aberto (Q_2, Q_4) , a receita total R excede o custo total C , e, assim, π é positivo. Mas, nos intervalos $[0, Q_2)$ e $(Q_4, Q_5]$, onde Q_5 representa o limite superior da capacidade de produção da empresa, π é negativo. Esse fato está refletido na Figura 9.7b, onde a curva de lucro – obtida traçando-se a distância vertical entre as curvas R e C para cada nível de produção – está acima do eixo horizontal somente no intervalo (Q_2, Q_4) .

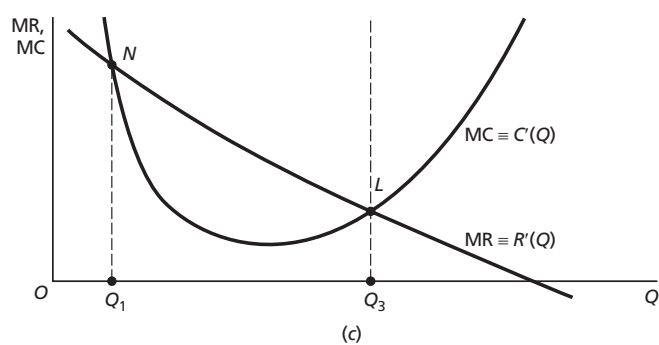
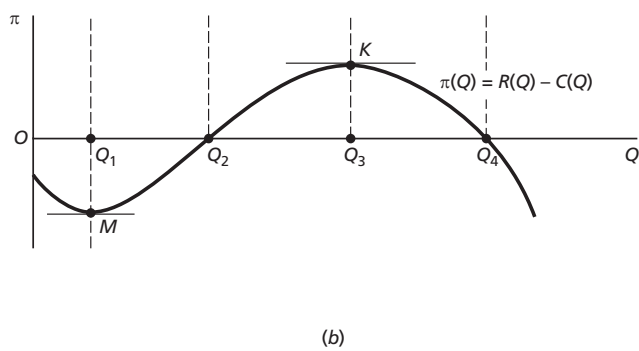
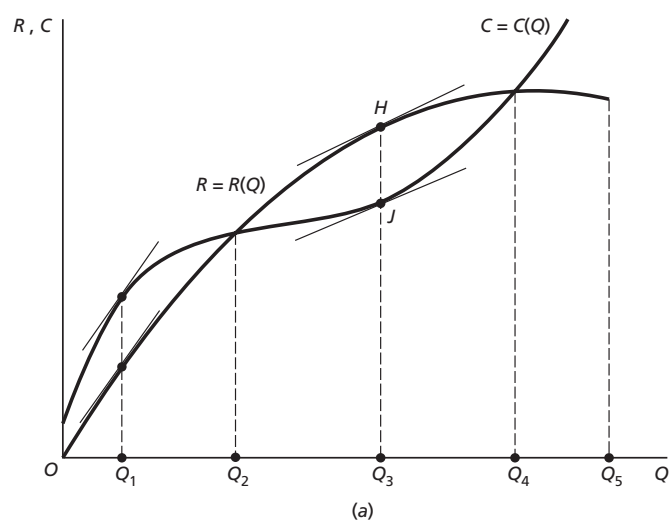
Quando estabelecemos $d\pi/dQ = 0$, segundo a condição de primeira ordem, nossa intenção é localizar o ponto de pico K sobre a curva de lucro na produção Q_3 , onde a inclinação da curva é zero. Todavia, o ponto de mínimo relativo M (produção Q_1) também aparecerá como um candidato, porque também cumpre o requisito de inclinação zero. Logo adiante recorreremos à condição de segunda ordem para eliminar o tipo “errado” de extremo.

A condição de primeira ordem $d\pi/dQ = 0$ é equivalente à condição $R'(Q) = C'(Q)$. Na Figura 9.7a, o nível de produção Q_3 satisfaz essa condição, porque as curvas R e C realmente têm a mesma inclinação em Q_3 (as retas tangentes às curvas, traçadas em H e J são paralelas). O mesmo é válido para a produção Q_1 . Uma vez que as inclinações iguais de R e C significam que MR e MC são iguais, as produções Q_3 e Q_1 devem estar, obviamente, onde as curvas MR e MC se interceptam, como ilustrado na Figura 9.7c.

Como a condição de segunda ordem entra em cena? Em primeiro lugar, vamos examinar a Figura 9.7b. No ponto K , a derivada segunda da função π (excetuando-se o caso excepcional do valor zero) terá um valor negativo, $\pi''(Q_3) < 0$, porque a curva tem a forma de U invertido ao redor de K ; isso significa que Q_3 maximizará o lucro. No ponto M , por outro lado, é de se esperar que $\pi''(Q_1) > 0$; assim, por sua vez, Q_1 provê um mínimo relativo para π . É claro que a condição suficiente de segunda ordem para um máximo pode ser enunciada alternativamente como $R''(Q) < C''(Q)$, isto é, que a inclinação da curva MR tem de ser menor que a inclinação da curva MC . Na Figura 9.7c, fica imediatamente aparente que a produção Q_3 satisfaz essa condição, já que a inclinação de MR é negativa enquanto a de MC é positiva no ponto L . Mas a produção Q_1 viola essa condição porque as inclinações de ambas, MC e MR , são negativas e a de MR é *numericamente menor* que a de MC no ponto N , o que implica, por sua vez, que $R''(Q_1)$ é *maior* que $C''(Q_1)$.

Portanto, na verdade, a produção Q_1 também viola a condição *necessária* de segunda ordem para um máximo relativo, mas satisfaz a condição *suficiente* de segunda ordem para um mínimo relativo.

FIGURA 9.7



EXEMPLO 3 Considere as funções $R(Q)$ e $C(Q)$ dadas por

$$R(Q) = 1200Q - 2Q^2$$

$$C(Q) = Q^3 - 61,25Q^2 + 1528,5Q + 2000$$

Então, a função lucro é

$$\pi(Q) = -Q^3 + 59,25Q^2 - 328,5Q - 2000$$

onde R , C e π estão em unidades de dólar e Q está em unidades de (digamos) toneladas por semana. Essa função lucro tem dois valores críticos, $Q = 3$ e $Q = 36,5$, porque

$$\frac{d\pi}{dQ} = -3Q^2 + 118,5Q - 328,5 = 0 \quad \text{quando } Q = \begin{cases} 3 \\ 36,5 \end{cases}$$

Mas, visto que a derivada segunda é

$$\frac{d^2\pi}{dQ^2} = -6Q + 118,5 \quad \begin{cases} > 0 & \text{quando } Q = 3 \\ < 0 & \text{quando } Q = 36,5 \end{cases}$$

a produção que maximiza o lucro é $Q^* = 36,5$ (toneladas por semana). (A outra produção minimiza o lucro). Substituindo Q^* na função lucro, podemos constatar que o lucro maximizado é $\pi^* = \pi(36,5) = 16.318,44$ (dólares por semana).

Como uma alternativa à abordagem precedente, podemos primeiramente achar as funções MR e MC e então igualar as duas, isto é, achar sua interseção. Visto que

$$R'(Q) = 1200 - 4Q$$

$$C'(Q) = 3Q^2 - 122,5Q + 1528,5$$

igualar as duas funções resultará em uma equação quadrática idêntica a $d\pi/dQ = 0$, que resultou nos dois valores críticos de Q já citados.

Coeficientes de uma função de custo total cúbica

No Exemplo 3, é utilizada uma função cúbica para representar a função de custo total. A curva tradicional de custo total $C = C(Q)$, como ilustrada na Figura 9.7a, deve conter uma mudança de curvatura formando um segmento côncavo (custo marginal decrescente) e um segmento convexo subsequente (custo marginal crescente). Posto que o gráfico de uma função cúbica sempre contém exatamente uma mudança de curvatura, como ilustrado na Figura 9.4, ele deve se adaptar bem àquele papel. Contudo, a Figura 9.4 nos alerta imediatamente para um problema: a função cúbica pode ter um segmento de inclinação descendente em seu gráfico, ao passo que, para ter sentido econômico, a função de custo total deve ter inclinação ascendente em toda a sua extensão (uma produção maior sempre acarreta um custo total maior). Por conseguinte, se quisermos usar uma função de custo total cúbica, tal como

$$C = C(Q) = aQ^3 + bQ^2 + cQ + d \quad (9.3)$$

é essencial impor restrições adequadas aos parâmetros de modo a impedir que a curva C vire para baixo.

Um modo equivalente de enunciar esse requisito é que a função MC deve ser positiva em toda a sua extensão, e isso pode ser assegurado somente se o *mínimo absoluto* da função MC resultar positivo. Diferenciando (9.3) em relação a Q , obtemos a função MC

$$MC = C'(Q) = 3aQ^2 + 2bQ + c \quad (9.4)$$

cujo gráfico, por ser uma função quadrática, é uma parábola, como na Figura 9.7c. Para que a curva MC permaneça positiva (acima do eixo horizontal) em toda a sua extensão, é necessário que a parábola tenha a forma de U (caso seja um U invertido, a curva fatalmente se estenderá para o segundo quadrante). Daí, o coeficiente do termo Q^2 em (9.4) tem de ser positivo, isto é, devemos impor a restrição $a > 0$. Contudo, essa restrição não é, de modo algum, suficiente, porque, ainda assim, o valor mínimo de uma curva MC em U – vamos denominá-lo $MC_{\min}$ (um mínimo relativo que, por acaso, também é um mínimo absoluto) – pode ocorrer abaixo do eixo horizontal. Por isso, devemos achar $MC_{\min}$ em seguida e determinar as restrições de parâmetros que a tornariam positiva.

Conforme o que sabemos do extremo relativo, o mínimo de MC ocorrerá onde

$$\frac{d}{dQ} MC = 6aQ + 2b = 0$$

O nível de produção que satisfaz essa condição de primeira ordem é

$$Q^* = \frac{-2b}{6a} = \frac{-b}{3a}$$

Isso minimiza (em vez de maximizar) MC, porque a derivada segunda $d^2(MC)/dQ^2 = 6a$ é, sem dúvida, positiva, em vista da restrição $a > 0$. Conhecer Q^* agora nos habilita a calcular $MC_{\min}$, mas dela podemos primeiramente inferir o sinal do coeficiente b . Considerando que são excluídos níveis negativos de produção, vemos que b nunca pode ser positivo (dado $a > 0$). Além disso, posto que supõe-se que a lei dos retornos decrescentes vale em um nível positivo de produção (isto é, supõe-se que MC tem um segmento inicial decrescente), Q^* deve ser positiva (em vez de zero). Por consequência, devemos impor a restrição $b < 0$.

Agora, basta tão somente substituir a produção Q^* que minimiza MC em (9.4) para constatar que

$$MC_{\min} = 3a \left(\frac{-b}{3a} \right)^2 + 2b \frac{-b}{3a} + c = \frac{3ac - b^2}{3a}$$

Assim, para garantir a positividade de $MC_{\min}$, devemos impor a restrição[†] $b^2 < 3ac$. A propósito, essa última restrição, na verdade, implica também a restrição $c > 0$. (Por quê?)

A discussão precedente envolveu os três parâmetros a , b e c . E o outro parâmetro, d ? A resposta é que também há necessidade de uma restrição sobre d , mas ela não tem nada a ver com o problema de manter MC positiva. Se considerarmos $Q=0$ em (9.3), constatamos que $C(0) = d$. Assim, o papel de d é apenas determinar a interseção vertical da curva C, sem nenhuma interferência sobre sua inclinação. Uma vez que o significado econômico de d é o custo fixo de uma empresa, a restrição adequada (no contexto de curto prazo) seria $d > 0$.

Em suma, os coeficientes da função de custo total (9.3) devem ser restringidos da seguinte maneira (admitindo-se o contexto de curto prazo):

$$a, c, d > 0 \quad b < 0 \quad b^2 < 3ac \quad (9.5)$$

Como você pode verificar de pronto, a função $C(Q)$ no Exemplo 3 satisfaz (9.5).

Curva de receita marginal de inclinação crescente

A Figura 9.7c mostra a curva de receita marginal com inclinação descendente em toda a sua extensão. Evidentemente, é assim que a curva MR é tradicionalmente desenhada para uma empresa sob condições de concorrência imperfeita. Contudo, a possibilidade de a inclinação da curva

[†] Essa restrição também pode ser obtida pelo método de *completar o quadrado*. A função MC pode ser transformada sucessivamente da seguinte maneira:

$$\begin{aligned} MC &= 3aQ^2 + 2bQ + c \\ &= \left(3aQ^2 + 2bQ + \frac{b^2}{3a} \right) - \frac{b^2}{3a} + c \\ &= \left(\sqrt{3a}Q + \sqrt{\frac{b^2}{3a}} \right)^2 + \frac{-b^2 + 3ac}{3a} \end{aligned}$$

Já que existe a possibilidade de a expressão ao quadrado ser zero, para garantir a positividade de MC temos de impor que $b^2 < 3ac$ sabendo que $a > 0$.

MR ser parcialmente, ou até mesmo totalmente, descendente não pode ser, de modo algum, descartada *a priori*.[†]

Dada uma função de receita média $AR = f(Q)$, a função de receita marginal pode ser expressa por

$$MR = f(Q) + Qf'(Q) \quad [\text{de (7.7)}]$$

Assim, a inclinação da curva MR pode ser determinada pela derivada

$$\frac{d}{dQ} MR = f'(Q) + f'(Q) + Qf''(Q) = 2f'(Q) + Qf''(Q)$$

Enquanto a inclinação da curva AR for descendente (como seria sob concorrência imperfeita), o termo $2f'(Q)$ é, com absoluta certeza, negativo. Mas o termo $Qf''(Q)$ pode ser negativo, zero ou positivo, dependendo do sinal da derivada segunda da função AR, isto é, dependendo de a curva AR ser estritamente côncava, linear ou estritamente convexa. Se a curva AR for estritamente convexa em toda a sua extensão (como ilustrado na Figura 7.2) ou ao longo de um segmento específico, existirá a possibilidade de o termo $Qf''(Q)$ (positivo) dominar o termo $2f'(Q)$ (negativo), fazendo com que a inclinação da curva MR seja total ou parcialmente ascendente.

EXEMPLO 4 Considere a função de receita média

$$AR = f(Q) = 8000 - 23Q + 1,1Q^2 - 0,018Q^3$$

Como se pode verificar (veja Exercício 9.4–7), essa função dá origem a uma curva AR de inclinação descendente, como deve ser para uma empresa sob concorrência imperfeita. Visto que

$$MR = f(Q) + Qf'(Q) = 8000 - 46Q + 3,3Q^2 - 0,072Q^3$$

deduz-se a inclinação de MR é

$$\frac{d}{dQ} MR = -46 + 6,6Q - 0,216Q^2$$

Como essa é uma função quadrática, e visto que o coeficiente de Q^2 é negativo, o gráfico de dMR/dQ em função de Q deve ser uma curva na forma de U invertido, como mostra a Figura 9.5a. Caso um segmento dessa curva esteja acima do eixo horizontal, a inclinação de MR assumirá valores positivos.

Fazendo $dMR/dQ = 0$ e aplicando a fórmula quadrática, constatamos que os dois zeros da função quadrática são $Q_1 = 10,76$ e $Q_2 = 19,79$ (aproximadamente). Isso significa que, para valores de Q no intervalo aberto (Q_1, Q_2) , a curva dMR/dQ está acima do eixo horizontal. Assim, a inclinação da curva de receita marginal é, de fato, positiva para níveis de produção entre Q_1 e Q_2 .

A presença de um segmento da curva MR com inclinação positiva tem implicações interessantes. Uma curva MR desse tipo pode produzir mais de uma interseção com a curva MC, satisfazendo a condição suficiente de segunda ordem para maximização do lucro. Conquanto todas essas interseções constituam ótimos locais, somente um deles é o ótimo global que a empresa está procurando.

EXERCÍCIO 9.4

1. Calcule os máximos e os mínimos relativos de y pelo teste da derivada segunda:

$$(a) y = -2x^2 + 8x + 25 \quad (c) y = \frac{1}{3}x^3 - 3x^2 + 5x + 3$$

$$(b) y = x^3 + 6x^2 + 9 \quad (d) y = \frac{2x}{1-2x} \quad \left(x \neq \frac{1}{2} \right)$$

[†] Esse ponto é destacado enfaticamente por John P. Formby, Layson, Stephen, e Smith, W. James. In: "The Law of Demand, Positive Sloping Marginal Revenue, and Multiple Profit Equilibria". *Economic Inquiry*, p. 303–311, abr. 1982.

2. O sr. Jardineiro quer demarcar um canteiro retangular de flores utilizando uma das paredes de sua casa como um lado do retângulo. Os outros três lados serão demarcados com alambrado de metal, do qual ele possui apenas 20 metros. Quais são o comprimento L e a largura W do retângulo que lhe dariam a maior área de plantio possível? Como ter certeza de que sua resposta é a maior área, e não a menor?
3. Uma empresa tem as seguintes funções, custo total e, demanda:

$$C = \frac{1}{3}Q^3 - 7Q^2 + 111Q + 50$$

$$Q = 100 - P$$
 - (a) A função de custo total satisfaz as restrições sobre os coeficientes dadas em (9.5)?
 - (b) Escreva a função de receita total R em termos de Q .
 - (c) Formule a função de lucro total π em termos de Q .
 - (d) Encontre o nível de produção Q^* que maximiza o lucro.
 - (e) Qual é o lucro máximo?
4. Se o coeficiente b em (9.3) fosse nulo, o que aconteceria com as curvas de custo marginal e de custo total?
5. Uma função de lucro quadrática $\pi(Q) = hQ^2 + jQ + k$ será usada para refletir as seguintes premissas:
 - (a) Se nada for produzido, o lucro será negativo (por causa dos custos fixos).
 - (b) A função de lucro é estritamente côncava.
 - (c) O lucro máximo ocorre em um nível positivo de produção Q^* .
 - (d) Que restrições sobre os parâmetros são necessárias?
6. Uma empresa puramente competitiva tem um único insumo variável L (mão-de-obra), cuja taxa salarial é W_0 por período. Os insumos fixos custam à empresa um total de F dólares por período. O preço do produto é P_0 .
 - (a) Escreva a função de produção, a função de receita, a função de custo e a função de lucro da empresa.
 - (b) Qual é a condição de primeira ordem para a maximização do lucro? Dê uma interpretação econômica a essa condição.
 - (c) Quais circunstâncias econômicas garantiriam que o lucro fosse maximizado, em vez de minimizado?
7. Use o seguinte procedimento para verificar se a inclinação da curva AR no Exemplo 4 é negativa:
 - (a) Denote a inclinação de AR por S . Escreva uma expressão para S .
 - (b) Calcule o valor máximo de S , $S_{\max}$, usando o teste da derivada segunda.
 - (c) Então, deduza, pelo valor de $S_{\max}$, que a inclinação da curva AR é negativa em toda a sua extensão.

9.5 Série de Maclaurin e série de Taylor

Chegou a hora de desenvolver um teste para extremos relativos que pode ser aplicado mesmo quando a derivada segunda tem um valor zero no ponto estacionário. Antes de mais nada, entretanto, é necessário discutir o que denominamos a expansão de uma função $y = f(x)$ para o que são conhecidas respectivamente como uma *série de Maclaurin* (expansão ao redor do ponto $x = 0$) e uma *série de Taylor* (expansão ao redor de qualquer ponto $x = x_0$).

Expandir uma função $y = f(x)$ ao redor de um ponto x_0 significa, no presente contexto, transformar aquela função em uma forma *polinomial*, na qual os coeficientes dos vários termos são expressos em termos dos valores das derivadas $f'(x_0)$, $f''(x_0)$ etc. – todos calculados no ponto de expansão x_0 . Na série de Maclaurin, eles serão calculados em $x = 0$; teremos, assim, $f'(0)$, $f''(0)$ etc. nos coeficientes. O resultado da expansão é uma *série de potências* porque, sendo um polinômio, ela consiste em uma soma de funções de potência.

Série de Maclaurin de uma função polinomial

Vamos considerar primeiro a expansão de uma função *polinomial* de n -ésimo grau,

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots + a_nx^n \quad (9.6)$$

para um polinômio equivalente de n -ésimo grau no qual os coeficientes (a_0, a_1 etc.) são expressos em termos dos valores das derivadas $f'(0), f''(0)$ etc. Visto que isso envolve a transformação de um polinômio em outro do mesmo grau, talvez pareça um exercício estéril e sem propósito, mas, na verdade, servirá para esclarecer muito bem a idéia de expansão.

Uma vez que a série de potências após a expansão envolverá as derivadas de várias ordens da função f , em primeiro lugar, vamos achar essas derivadas. Por diferenciação sucessiva de (9.6), podemos obter as derivadas como se segue:

$$\begin{aligned} f'(x) &= a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \cdots + na_nx^{n-1} \\ f''(x) &= 2a_2 + 3(2)a_3x + 4(3)a_4x^2 + \cdots + n(n-1)a_nx^{n-2} \\ f'''(x) &= 3(2)a_3 + 4(3)(2)a_4x + \cdots + n(n-1)(n-2)a_nx^{n-3} \\ f^{(4)}(x) &= 4(3)(2)a_4 + 5(4)(3)(2)a_5x + \cdots + n(n-1)(n-2)(n-3)a_nx^{n-4} \\ &\vdots \\ f^{(n)}(x) &= n(n-1)(n-2)(n-3) \cdots (3)(2)(1)a_n \end{aligned}$$

Note que, a cada diferenciação sucessiva, o número de termos é reduzido um de cada vez – a constante aditiva da frente desaparece – até que, na n -ésima derivada, resta apenas um único termo de produto (um termo constante). Essas derivadas podem ser calculadas em vários valores de x ; aqui, nós as calcularemos em $x=0$, o que resulta no descarte de todos os termos que envolvem x . Ficamos, então, com os seguintes valores de derivadas excepcionalmente claros:

$$\begin{aligned} f'(0) &= a_1 & f''(0) &= 2a_2 & f'''(0) &= 3(2)a_3 & f^{(4)}(0) &= 4(3)(2)a_4 \\ \cdots & & f^{(n)}(0) &= n(n-1)(n-2)(n-3) \cdots (3)(2)(1)a_n \end{aligned} \quad (9.7)$$

Se agora adotarmos um símbolo abreviado $n!$ (lê-se “fatorial de n ”), definido como

$$n! \equiv n(n-1)(n-2) \cdots (3)(2)(1) \quad (n = \text{um inteiro positivo})$$

de modo que, por exemplo, $2! = 2 \times 1 = 2$ e $3! = 3 \times 2 \times 1 = 6$ etc. (sendo $0!$ definido como igual a 1), então o resultado em (9.7) pode ser reescrito como

$$a_1 = \frac{f'(0)}{1!} \quad a_2 = \frac{f''(0)}{2!} \quad a_3 = \frac{f'''(0)}{3!} \quad a_4 = \frac{f^{(4)}(0)}{4!} \quad \cdots \quad a_n = \frac{f^{(n)}(0)}{n!}$$

Substituindo-os em (9.6) e utilizando o fato óbvio de que $f(0) = a_0$, agora podemos expressar a função dada $f(x)$ como um novo polinômio do mesmo grau, mas equivalente, no qual os coeficientes são expressos em termos das derivadas calculadas em $x=0$:[†]

$$\begin{aligned} f(x) &= \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 \\ &\quad + \cdots + \frac{f^{(n)}(0)}{n!}x^n \quad [\text{fórmula de Maclaurin}] \end{aligned} \quad (9.8)$$

[†] Visto que $0! = 1$ e $1! = 1$, os dois primeiros termos à direita do sinal de igualdade em (9.8) podem ser escritos mais simplesmente como $f(0)$ e $f'(0)x$, respectivamente. Incluímos aqui os denominadores $0!$ e $1!$ para chamar a atenção para a simetria entre os vários termos na expansão.

Esse novo polinômio, denominado série de Maclaurin da função polinomial $f(x)$, representa a expansão da função $f(x)$ ao redor de zero ($x=0$). Note que o ponto de expansão (aqui, 0), é simplesmente o valor de x que será usado para calcular $f(x)$ e todas as suas derivadas.

EXEMPLO 1 Encontre a série de Maclaurin para a função

$$f(x) = 2 + 4x + 3x^2 \quad (9.9)$$

Essa função tem as derivadas

$$\begin{aligned} f'(x) &= 4 + 6x \\ f''(x) &= 6 \end{aligned} \quad \text{de modo que} \quad \begin{cases} f'(0) = 4 \\ f''(0) = 6 \end{cases}$$

Assim, a série de Maclaurin é

$$\begin{aligned} f(x) &= f(0) + f'(0)x + \frac{f''(0)}{2}x^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

A linha anterior prova que a série de Maclaurin de fato representa corretamente a função dada.

Série de Taylor de uma função polinomial

Em termos mais gerais, a função polinomial em (9.6) pode ser expandida ao redor de qualquer ponto x_0 , não necessariamente zero. Por questão de simplicidade, explicaremos isso por meio da função quadrática específica em (9.9) e generalizaremos o resultado mais adiante.

Para a finalidade de expansão ao redor de um ponto específico x_0 , podemos primeiro interpretar qualquer valor dado de x como um *desvio* em relação a x_0 . Mais especificamente, seja $x = x_0 + \delta$, onde δ representa o desvio em relação ao valor x_0 . Por essa interpretação, a função dada (9.9) e suas derivadas agora se tornam

$$\begin{aligned} f(x) &= 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \\ f'(x) &= 4 + 6(x_0 + \delta) \\ f''(x) &= 6 \end{aligned} \quad (9.10)$$

Sabemos que a expressão $(x_0 + \delta) = x$ é uma variável na função, mas, visto que no presente contexto x_0 é um número *fixo* (escolhido), somente δ pode ser considerada adequadamente como uma variável em (9.10). Por consequência, $f(x)$ é, de fato, uma função de δ , digamos, $g(\delta)$:

$$g(\delta) = 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \quad [\equiv f(x)]$$

com derivadas

$$\begin{aligned} g'(\delta) &= 4 + 6(x_0 + \delta) \quad [\equiv f'(x)] \\ g''(\delta) &= 6 \quad [\equiv f''(x)] \end{aligned}$$

Já sabemos como expandir $g(\delta)$ ao redor de zero ($\delta = 0$). Conforme (9.8), tal expansão dará a seguinte série de Maclaurin:

$$g(\delta) = \frac{g(0)}{0!} + \frac{g'(0)}{1!}\delta + \frac{g''(0)}{2!}\delta^2 \quad (9.11)$$

Mas, visto que $x = x_0 + \delta$, o fato de que $\delta = 0$ implica $x = x_0$; daí, com base na identidade $g(\delta) \equiv f(x)$, podemos escrever, para o caso de $\delta = 0$:

$$g(0) = f(x_0) \quad g'(0) = f'(x_0) \quad g''(0) = f''(x_0)$$

Substituindo esses valores em (9.11), verificamos que o resultado representa a expansão de $f(x)$ ao redor do ponto x_0 , porque o coeficiente agora envolve as derivadas $f'(x_0)$, $f''(x_0)$ etc., todas calculadas em $x = x_0$:

$$f(x) [=g(\delta)] = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 \quad (9.12)$$

Você deve comparar esse resultado – o polinômio de Taylor de $f(x)$ – com o polinômio de Maclaurin de $g(\delta)$ em (9.11).

Uma vez que, para a função específica em consideração, (9.9), temos

$$f(x_0) = 2 + 4x_0 + 3x_0^2 \quad f'(x_0) = 4 + 6x_0 \quad f''(x_0) = 6$$

o polinômio de Taylor em (9.12) torna-se

$$\begin{aligned} f(x) &= 2 + 4x_0 + 3x_0^2 + (4 + 6x_0)(x - x_0) + \frac{6}{2}(x - x_0)^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

Isso prova que o polinômio de Taylor de fato representa corretamente a função dada.

A fórmula de expansão em (9.12) pode ser generalizada para se aplicar ao polinômio de n -ésimo grau de (9.6). A fórmula generalizada é

$$\begin{aligned} f(x) &= \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots \\ &\quad + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \quad [\text{fórmula de Taylor}] \end{aligned} \quad (9.13)$$

A diferença entre essa fórmula e a de Maclaurin em (9.8) é apenas na substituição de zero por x_0 como o ponto de expansão e na substituição de x pela expressão $(x - x_0)$. O que (9.13) nos diz é que, dado um polinômio de n -ésimo grau $f(x)$, se fizemos $x = 7$ (digamos) nos termos à direita de (9.13), selecionarmos um número arbitrário x_0 , e então calculamos e adicionarmos esses termos, teremos exatamente $f(7)$ – o valor de $f(x)$ em $x = 7$.

EXEMPLO 2 Tomando $x_0 = 3$ como o ponto de expansão, podemos reescrever (9.6) de modo equivalente como

$$f(x) = f(3) + f'(3)(x - 3) + \frac{f''(3)}{2}(x - 3)^2 + \dots + \frac{f^{(n)}(3)}{n!}(x - 3)^n$$

Expansão de uma função arbitrária

Até aqui, mostramos como uma função polinomial de n -ésimo grau pode ser expressa sob a forma de um outro polinômio de n -ésimo grau, equivalente. Na verdade, também é possível expressar qualquer função *arbitrária* $\phi(x)$ – uma função que não seja necessariamente um polinômio – em uma forma polinomial similar a (9.13), contanto que $\phi(x)$ tenha derivadas finitas, contínuas até a ordem desejada no ponto de expansão x_0 .

De acordo com uma proposição matemática conhecida como *teorema de Taylor*, dada uma função arbitrária $\phi(x)$, se conhecermos o valor da função em $x = x_0$ [isto é, $\phi(x_0)$] e os valores de suas derivadas em x_0 [isto é, $\phi'(x_0)$, $\phi''(x_0)$ etc.], então essa função pode ser expandida ao redor do ponto x_0 como se segue ($n =$ um número inteiro positivo fixo escolhido arbitrariamente):

$$\begin{aligned} \phi(x) &= \left[\frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x-x_0) + \frac{\phi''(x_0)}{2!}(x-x_0)^2 \right. \\ &\quad \left. + \cdots + \frac{\phi^{(n)}(x_0)}{n!}(x-x_0)^n \right] + R_n \\ &\equiv P_n + R_n \text{ [fórmula de Taylor com resto]} \end{aligned} \quad (9.14)$$

onde P_n representa (entre colchetes) um polinômio de n -ésimo grau [os primeiros $(n+1)$ termos à direita] e R_n denota um *resto* que será explicado na página 236.[†] A presença de R_n é o que distingue (9.14) da fórmula de Taylor (9.13) e, por essa razão, (9.14) é denominada *fórmula de Taylor com resto*. A forma do polinômio P_n e a grandeza do resto R_n dependerão do valor de n que escolhermos. Quanto maior o n , mais termos haverá em P_n ; de acordo com isso, em geral R_n assumirá um valor diferente para cada n diferente. Esse fato explica a necessidade do índice n nesses dois símbolos. Como recurso mnemônico, podemos identificar n como a ordem da derivada mais alta em P_n . (No caso especial de $n=0$, não aparecerá absolutamente nenhuma derivada em P_n .)

A aparição de R_n em (9.14) deve-se ao fato de agora estarmos tratando com uma função arbitrária ϕ que nem sempre pode ser transformada *exatamente* na forma polinomial mostrada em (9.13), mas apenas aproximadamente. Por conseguinte, um termo de resto é incluído como um suplemento da parte P_n , para representar a discrepância entre $\phi(x)$ e P_n . Assim, P_n constitui uma aproximação polinomial de $\phi(x)$, sendo o termo R_n uma medida do erro de aproximação. Se escolhermos $n=1$, por exemplo, temos

$$\phi(x) = [\phi(x_0) + \phi'(x_0)(x-x_0)] + R_1 = P_1 + R_1$$

onde P_1 consiste em $n+1=2$ termos e constitui uma aproximação *linear* de $\phi(x)$. Se escolhermos $n=2$, aparecerá um termo de segundo grau, de modo que

$$\phi(x) = \left[\phi(x_0) + \phi'(x_0)(x-x_0) + \frac{\phi''(x_0)}{2!}(x-x_0)^2 \right] + R_2 = P_2 + R_2$$

onde P_2 , consistindo em $n+1=3$ termos, é uma aproximação *quadrática* de $\phi(x)$. E assim por diante. O fato de podermos criar aproximações polinomiais para qualquer função arbitrária (contanto que tenha derivadas finitas, contínuas) é de grande significância prática. Funções polinomiais – mesmo as de graus mais altos – são relativamente fáceis para trabalhar e, se puderem servir como boas aproximações para algumas funções difíceis, ganha-se muito em conveniência, como ilustrarão os dois próximos exemplos.

É preciso ressaltar que a função arbitrária $\phi(x)$ poderia, obviamente, abranger o polinômio de n -ésimo grau de (9.6) como um caso especial. Para este último caso, se a expansão for para um outro polinômio de n -ésimo grau, o resultado de (9.13) se aplicará exatamente; ou, em outras palavras, podemos usar o resultado em (9.14), com $R_n \equiv 0$. Contudo, se o polinômio de n -ésimo grau dado, $f(x)$, for expandido para um polinômio de *menor* grau, então este último pode ser considerado apenas uma aproximação de $f(x)$, e deve aparecer um resto; nesse caso, o resultado em (9.14) se aplica, porém com um resto diferente de zero. Assim, a fórmula de Taylor na forma de (9.14) é perfeitamente geral.

EXEMPLO 3 Expanda a função não-polinomial

$$\phi(x) = \frac{1}{1+x}$$

ao redor do ponto $x_0 = 1$, com $n = 4$. Precisaremos das quatro primeiras derivadas de $\phi(x)$, que são

[†] O símbolo R_n (resto) não deve ser confundido com o símbolo R^n (espaço de n dimensões).

$$\begin{aligned}
 \phi'(x) &= -(1+x)^{-2} & \text{de modo que} & & \phi'(1) &= -(2)^{-2} = -\frac{1}{4} \\
 \phi''(x) &= 2(1+x)^{-3} & & & \phi''(1) &= 2(2)^{-3} = \frac{1}{4} \\
 \phi'''(x) &= -6(1+x)^{-4} & & & \phi'''(1) &= -6(2)^{-4} = -\frac{3}{8} \\
 \phi^{(4)}(x) &= 24(1+x)^{-5} & & & \phi^{(4)}(1) &= 24(2)^{-5} = \frac{3}{4}
 \end{aligned}$$

Vemos, também, que $\phi(1) = \frac{1}{2}$. Assim, estabelecendo $x_0 = 1$ em (9.14) e utilizando as derivadas obtidas, chegamos à seguinte série de Taylor com resto:

$$\begin{aligned}
 \phi(x) &= \frac{1}{2} - \frac{1}{4}(x-1) + \frac{1}{8}(x-1)^2 - \frac{1}{16}(x-1)^3 + \frac{1}{32}(x-1)^4 + R_4 \\
 &= \frac{31}{32} - \frac{13}{16}x + \frac{1}{2}x^2 - \frac{3}{16}x^3 + \frac{1}{32}x^4 + R_4
 \end{aligned}$$

É claro que aqui também é possível escolher $x_0 = 0$ como o ponto de expansão. Nesse caso, com x_0 igual a zero em (9.14), a expansão resultará em uma *série de Maclaurin com resto*.

EXEMPLO 4

Expanda a função quadrática

$$\phi(x) = 5 + 2x + x^2$$

ao redor de $x_0 = 1$, com $n = 1$. Essa função, assim como (9.9) no Exemplo 1, é um polinômio de segundo grau. Mas, visto que $n = 1$, nossa tarefa será expandi-lo para um polinômio de *primeiro* grau, isto é, achar uma aproximação linear da função quadrática dada; assim, é certo que aparecerá um termo de resto. Por essa razão, $\phi(x)$ deve ser vista como uma função “arbitrária” para o propósito dessa expansão de Taylor.

Para efetuar essa expansão, precisamos somente da derivada primeira $\phi'(x) = 2 + 2x$. Calculada em $x_0 = 1$, a função dada e sua derivada resultam em

$$\phi(x_0) = \phi(1) = 8 \quad \phi'(x_0) = \phi'(1) = 4$$

Assim, a fórmula de Taylor com resto nos dá

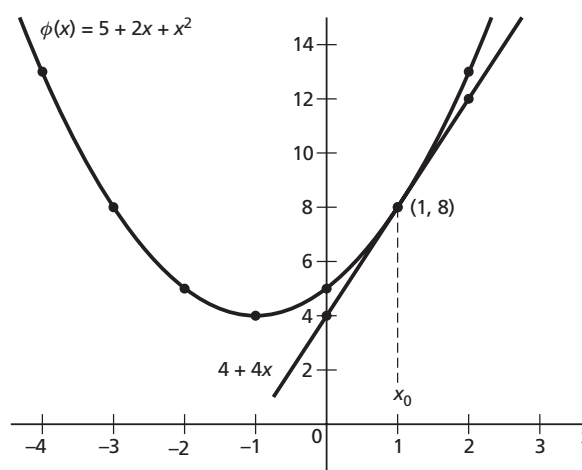
$$\begin{aligned}
 \phi(x) &= \phi(x_0) + \phi'(x_0)(x - x_0) + R_1 \\
 &= 8 + 4(x - 1) + R_1 = 4 + 4x + R_1
 \end{aligned}$$

onde o termo $(4 + 4x)$ é uma aproximação linear e o termo R_1 representa o erro de aproximação.

Na Figura 9.8, o gráfico de $\phi(x)$ é uma parábola, e o gráfico de sua aproximação linear é uma reta tangente à curva $\phi(x)$ no ponto $(1, 8)$. A ocorrência do ponto de tangência em $x = 1$ não é coincidência; ao contrário, é a consequência direta do fato de o ponto de expansão ter sido estabelecido naquele valor específico de x . Isso sugere que, quando uma função arbitrária $\phi(x)$ é aproximada por um polinômio, este último dará o valor exato de $\phi(x)$ *no* (e somente *no*) ponto de expansão, com erro de aproximação zero ($R_1 = 0$). Em todos os outros lugares, R_1 é estritamente diferente de zero e, na verdade, mostra erros de aproximação cada vez maiores à medida que tentamos aproximar $\phi(x)$ para valores de x cada vez mais longe do ponto de expansão x_0 . Assim, quando tentamos aproximar qualquer função $\phi(x)$ por um polinômio, se estivermos mais interessados em obter uma aproximação precisa na vizinhança de um valor específico de x , digamos, x_0 , então temos de escolher x_0 como o ponto de expansão.

A construção da Figura 9.8 nos lembra muito a da Figura 8.1. De fato, ambas as figuras tratam de “aproximações”. Mas há uma diferença no escopo da aproximação. Na Figura 8.1, tentamos aproximar Δy pela diferencial dy com a ajuda de uma reta tangente traçada em x_0 , um dado valor de partida de x . Na Figura 9.8, por outro lado, nosso alvo é mais amplo – aproximar toda a curva por uma reta específica, isto é, aproximar a altura da curva em qualquer valor de x , digamos, x_1 , pela altura correspondente da reta em x_1 . Note que, em ambos os casos, o erro de aproximação varia com o valor de x . Na Figura 8.1, o erro (a diferença entre dy e Δy) fica menor à medida que Δx fica menor, ou à medida que x fica mais próximo de x_0 , ponto em que a reta tangente é traçada. Na Figura 9.8, o erro (a discrepância vertical entre a reta e a curva) fica menor à medida que x se aproxima de x_0 , o ponto de expansão escolhido.

FIGURA 9.8



Forma de Lagrange para o resto

Agora precisamos examinar melhor o termo de resto. De acordo com a *forma de Lagrange para o resto*, podemos expressar R_n como

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!} (x - x_0)^{n+1} \quad (9.15)$$

em que p é algum número entre x (o ponto onde desejamos calcular a função arbitrária ϕ) e x_0 (o ponto onde expandimos a função ϕ). Note que essa expressão é muito parecida com o termo que resultaria logicamente do último termo em P_n em (9.14), exceto que, aqui, a derivada envolvida deve ser avaliada em um ponto p , em vez de x_0 . Visto que, infelizmente, não há nenhuma especificação para o ponto p , essa fórmula não nos habilita realmente a calcular R_n ; não obstante, ela não deixa de ter grande significância analítica. Por conseguinte, vamos ilustrar seu significado graficamente, embora apenas para o caso simples de $n = 0$.

Quando $n = 0$, não aparecerá absolutamente nenhuma derivada na parte polinomial P_0 ; portanto, (9.14) é reduzida a

$$\phi(x) = P_0 + R_0 = \phi(x_0) + \phi'(p)(x - x_0)$$

ou

$$\phi(x) - \phi(x_0) = \phi'(p)(x - x_0)$$

Esse resultado, uma versão simples do *teorema do valor médio*, afirma que a diferença entre o valor da função ϕ em x_0 e qualquer outro valor de x pode ser expressa como o produto da diferença $(x - x_0)$ e da derivada ϕ' calculada em p (sendo p algum ponto entre x e x_0). Vamos examinar a Figura 9.9, que mostra o gráfico da função $\phi(x)$ como uma curva contínua com valores da derivada definidos em todos os pontos. Seja x_0 o ponto de expansão escolhido, e seja x qualquer ponto no eixo horizontal. Se tentarmos aproximar $\phi(x)$, a distância xB , por $\phi(x_0)$, a distância x_0A , estará envolvido um erro igual a $\phi(x) - \phi(x_0)$, a distância CB . O que o teorema do valor médio diz é que o erro CB – que constitui o valor do termo de resto R_0 na expansão – pode ser expresso como $\phi'(p)(x - x_0)$, onde p é algum ponto entre x e x_0 . Primeiro, localizamos um ponto D , entre os pontos A e B , tal que a reta tangente a D seja paralela à reta AB ; esse ponto D deve existir, já que a curva passa de A para B de uma maneira contínua e suave. Então, o resto será

$$\begin{aligned} R_0 = CB &= \frac{CB}{AC} AC = (\text{inclinação de } AB) \cdot AC \\ &= (\text{inclinação da tangente a } D) \cdot AC \\ &= (\text{inclinação da curva em } x = p) \cdot AC \\ &= \phi'(p)(x - x_0) \end{aligned}$$

4. Com base na fórmula de Taylor com o resto na forma de Lagrange [veja (9.14) e (9.15)], mostre que, no ponto de expansão ($x = x_0$), a série de Taylor sempre dará *exatamente* o valor da função naquele ponto, $\phi(x_0)$, e não uma mera aproximação.

9.6 Teste da derivada N -ésima para extremo relativo de uma função de uma só variável

A expansão de uma função em uma série de Taylor (ou de Maclaurin) é útil como um instrumento de aproximação na circunstância em que $R_n \rightarrow 0$ quando $n \rightarrow \infty$, mas o que nos preocupa agora é sua aplicação no desenvolvimento de um teste geral para um extremo relativo.

Expansão de Taylor e extremo relativo

Como etapa preparatória para essa tarefa, vamos redefinir um extremo relativo da seguinte maneira:

Uma função $f(x)$ atinge um valor máximo (mínimo) relativo em x_0 se $f(x) - f(x_0)$ for negativo (positivo) para valores de x na vizinhança imediata de x_0 , tanto à sua esquerda quanto à sua direita.

Isso pode ser esclarecido referindo-nos à Figura 9.10, onde x_1 é um valor de x à esquerda de x_0 , e x_2 é um valor de x à direita de x_0 . Na Figura 9.10a, $f(x_0)$ é um máximo relativo; assim, $f(x_0)$ é maior do que ambos, $f(x_1)$ e $f(x_2)$. Resumindo, $f(x) - f(x_0)$ é negativo para qualquer valor de x na vizinhança imediata de x_0 . O oposto é verdadeiro para a Figura 9.10b, onde $f(x_0)$ é um mínimo relativo e, assim, $f(x) - f(x_0) > 0$.

Supondo que $f(x)$ tem derivadas finitas, contínuas, até a ordem desejada no ponto $x = x_0$, a função $f(x)$ – não necessariamente polinomial – pode ser expandida ao redor do ponto x_0 como uma série de Taylor. Tomando como base (9.14) (após mudar devidamente ϕ para f), e usando a forma de Lagrange para o resto, podemos escrever

$$\begin{aligned} f(x) - f(x_0) = & f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots \\ & + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(p)}{(n+1)!}(x - x_0)^{n+1} \end{aligned} \quad (9.17)$$

Se o sinal da expressão $f(x) - f(x_0)$ puder ser determinado para valores de x à esquerda e à direita imediatas de x_0 , podemos, de pronto, chegar a uma conclusão se $f(x_0)$ é um extremo e, caso seja, se é um máximo ou um mínimo. Para isso, é necessário examinar a soma do lado direito de (9.17). Nessa soma há, ao todo, $(n+1)$ termos – n termos de P_n mais o resto, cujo grau é $(n+1)$ – e, assim, o número real de termos é indefinido, pois depende do valor que escolhermos para n . Contudo, escolhendo n adequadamente, sempre podemos ter certeza de que existirá somente um único termo à direita. Isso simplificará drasticamente a tarefa de avaliar o sinal de $f(x) - f(x_0)$ e determinará se $f(x_0)$ é um extremo e, se for, de que tipo.

Alguns casos específicos

Isso pode ficar ainda mais claro mediante algumas ilustrações específicas.

Caso 1

$$f'(x_0) \neq 0$$

Se a derivada primeira em x_0 não for diferente de zero, vamos escolher $n = 0$, de modo que o resto será de primeiro grau. Então, haverá somente $n + 1 = 1$ termo do lado direito, o que implica que lá estará somente o resto R_0 . Isto é, temos

$$f(x) - f(x_0) = \frac{f'(p)}{1!}(x - x_0) = f'(p)(x - x_0)$$

onde p é algum número entre x_0 e um valor de x na vizinhança imediata de x_0 . Note que, de acordo com isso, p deve estar muito, muito próximo de x_0 .

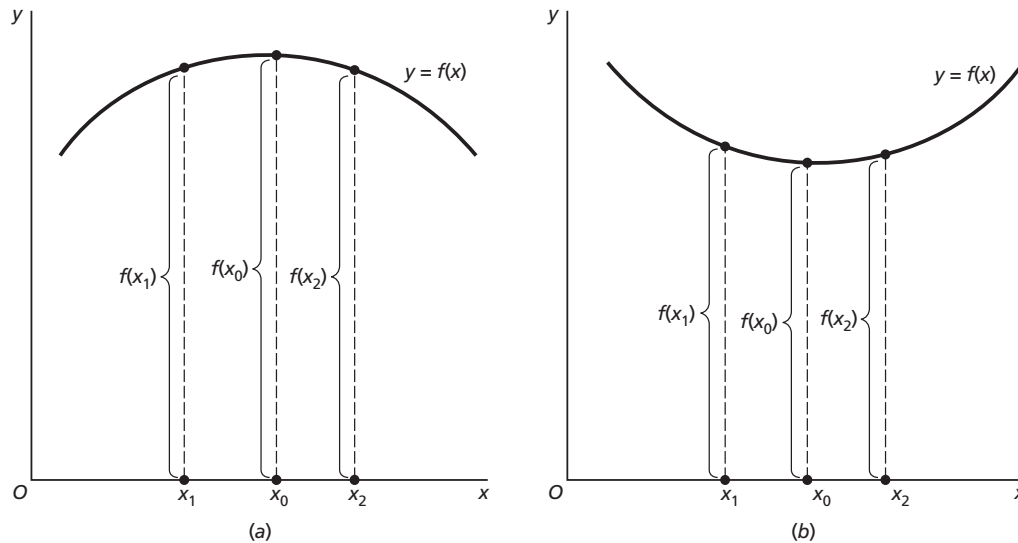


FIGURA 9.10

Qual é o sinal da expressão à direita? Por causa da continuidade da derivada, $f'(p)$ terá o mesmo sinal de $f'(x_0)$, visto que, como mencionado antes, p está muito, muito próximo de x_0 . No caso presente, $f'(p)$ deve ser diferente de zero; de fato, deve ser um número específico positivo ou negativo. Entretanto, e a parte $(x - x_0)$? Quando passamos da esquerda para a direita de x_0 , x passa de uma grandeza $x_1 < x_0$ para uma grandeza $x_2 > x_0$ (veja Figura 9.10). Por conseqüência, a expressão $(x - x_0)$ deve passar de negativa para positiva à medida que vamos adiante e $f(x) - f(x_0) = f'(p)(x - x_0)$ também deve mudar de sinal da esquerda para a direita de x_0 . Entretanto, isso viola nossa nova definição de um extremo relativo; de acordo com ela, não pode existir um extremo relativo em $f(x_0)$ quando $f'(x_0) \neq 0$ – um fato que já conhecemos bem.

Caso 2 $f'(x_0) = 0; f''(x_0) \neq 0$

Neste caso, escolha $n = 1$, de modo que o resto será de segundo grau. Então, haverá inicialmente $n + 1 = 2$ termos à direita. Mas um desses termos desaparecerá porque $f'(x_0) = 0$ e, então, ficaremos novamente com apenas um termo para calcular:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{f''(p)}{2!}(x - x_0)^2 \\ &= \frac{1}{2}f''(p)(x - x_0)^2 \quad [\text{porque } f'(x_0) = 0] \end{aligned}$$

Como antes, $f''(p)$ terá o mesmo sinal de $f''(x_0)$, um sinal que é especificado e invariável, ao passo que a parte $(x - x_0)^2$, sendo um quadrado, é invariavelmente positiva. Assim, a expressão $f(x) - f(x_0)$ deve assumir o mesmo sinal de $f''(x_0)$ e, de acordo com a definição anterior de extremo relativo, especificará

Um máximo relativo de $f(x)$ se $f''(x_0) < 0$
Um mínimo relativo de $f(x)$ se $f''(x_0) > 0$ [com $f'(x_0) = 0$]

Você reconhecerá que isso é o teste da derivada segunda apresentado anteriormente.

Caso 3 $f'(x_0) = f''(x_0) = 0$, mas $f'''(x_0) \neq 0$

Aqui, encontramos uma situação que o teste da derivada segunda é incapaz de manipular, pois $f''(x_0)$ agora é zero. Com a ajuda da série de Taylor, contudo, pode-se chegar a um resultado conclusivo sem dificuldade.

Vamos escolher $n = 2$; então, inicialmente, aparecerão três termos à direita. Porém, dois deles serão descartados porque $f'(x_0) = f''(x_0) = 0$, de modo que ficamos novamente com apenas um termo para calcular:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(p)(x - x_0)^3 \\ &= \frac{1}{6}f'''(p)(x - x_0)^3 \quad [\text{porque } f'(x_0) = 0, f''(x_0) = 0] \end{aligned}$$

Como antes, o sinal de $f'''(p)$ é idêntico ao sinal de $f'''(x_0)$ por causa da continuidade da derivada e porque p está muito próximo a x_0 . Mas a parte $(x - x_0)^3$ tem um sinal variável. Especificamente, visto que $(x - x_0)$ é negativo à esquerda de x_0 , $(x - x_0)^3$ também será; no entanto, à direita de x_0 , a parte $(x - x_0)^3$ será positiva. Assim, há uma mudança no sinal de $f(x) - f(x_0)$ quando passamos por x_0 , o que viola a definição de um extremo relativo. Contudo, sabemos que x_0 é um valor crítico [$f'(x_0) = 0$] e, assim, ele deve dar um ponto de inflexão, considerando que não dá um extremo relativo.

Caso 4 $f'(x_0) = f''(x_0) = \dots = f^{(N-1)}(x_0) = 0$, mas $f^{(N)}(x_0) \neq 0$

Este é um caso muito geral e, portanto, podemos derivar dele um resultado geral. Note que, aqui, os valores de todas as derivadas são nulos até chegarmos à N -ésima derivada.

De modo análogo aos três casos precedentes, a série de Taylor para o Caso 4 se reduzirá a

$$f(x) - f(x_0) = \frac{1}{N!}f^{(N)}(p)(x - x_0)^N$$

Mais uma vez, $f^{(N)}(p)$ tem o mesmo sinal que $f^{(N)}(x_0)$, que é invariável. O sinal da parte $(x - x_0)^N$, por outro lado, *variará* se N for *ímpar* (conforme Casos 1 e 3) e *permanecerá inalterado* (positivo) se N for *par* (conforme Caso 2). De acordo com isso, quando N for ímpar, $f(x) - f(x_0)$ mudará de sinal quando passarmos pelo ponto x_0 , violando, desse modo, a definição de um extremo relativo (o que significa que x_0 deve nos dar um ponto de inflexão na curva). Mas, quando N for par, $f(x) - f(x_0)$ não mudará de sinal da esquerda para a direita de x_0 , e isso estabelecerá o valor estacionário $f(x_0)$ como um máximo ou um mínimo relativo, dependendo de $f^{(N)}(x_0)$ ser negativa ou positiva.

Teste da N -ésima derivada

Por fim, então, podemos enunciar o seguinte teste geral.

Teste da N -ésima derivada para extremo relativo de uma função de uma só variável Se a derivada primeira de uma função $f(x)$ em x_0 for $f'(x_0) = 0$ e se o primeiro valor *diferente de zero de derivada* encontrado em x_0 por derivação sucessiva for o da N -ésima derivada, $f^{(N)}(x_0) \neq 0$, então o valor estacionário $f(x_0)$ será

- Um *máximo* relativo se N for um número par e $f^{(N)}(x_0) < 0$.
- Um *mínimo* relativo se N for um número par mas $f^{(N)}(x_0) > 0$.
- Um *ponto de inflexão* se N for ímpar.

Deve ficar claro, da declaração precedente, que o teste da N -ésima derivada pode funcionar se e somente se a função $f(x)$ tiver, mais cedo ou mais tarde, alguma derivada diferente de zero no valor crítico x_0 . Embora existam funções excepcionais que não satisfazem essa condi-

ção, a maioria das funções que provavelmente encontraremos de fato produzirão $f^{(N)}(x_0)$ diferente de zero por diferenciação sucessiva.[†] Assim, o teste deve provar ser útil na maioria dos casos.

EXEMPLO 1

Examine a função $y = (7 - x)^4$ em relação a seu extremo relativo. Visto que $f'(x) = -4(7 - x)^3$ é zero quando $x = 7$, tomamos $x = 7$ como o valor crítico para testar, sendo $y = 0$, o valor estacionário da função. Por derivação sucessiva (continuamos até encontrar um valor de derivada diferente de zero no ponto $x = 7$), obtemos

$$f''(x) = 12(7 - x)^2 \quad \text{de modo que} \quad f''(7) = 0$$

$$f'''(x) = -24(7 - x) \quad \quad \quad f'''(7) = 0$$

$$f^{(4)}(x) = 24 \quad \quad \quad f^{(4)}(7) = 24$$

Uma vez que 4 é um número par e visto que $f^{(4)}(7)$ é positiva, concluímos que o ponto $(7, 0)$ representa um mínimo relativo.

Como se pode verificar com facilidade, o gráfico dessa função é uma curva estritamente convexa. Considerando que a derivada segunda em $x = 7$ é zero (em vez de positiva), este exemplo serve para ilustrar nossa declaração anterior a respeito da derivada segunda e da curvatura de uma curva (Seção 9.3) no sentido de que, conquanto uma $f''(x)$ positiva para todo x implique uma $f(x)$ estritamente convexa, uma $f(x)$ estritamente convexa *não* implica uma $f''(x)$ positiva para todo x . Mais importante, ela também serve para ilustrar o fato de que, dada uma curva estritamente convexa (estritamente côncava), o extremo encontrado naquela curva deve ser um mínimo (máximo), porque tal extremo *ou* satisfará a condição suficiente de segunda ordem *ou*, se não, satisfará uma outra condição suficiente (de ordem mais alta) para um mínimo (máximo).

EXERCÍCIO 9.6

1. Encontre os valores estacionários das seguintes funções:

(a) $y = x^3$

(b) $y = -x^4$

(c) $y = x^6 + 5$

Determine, pelo teste da N -ésima derivada, se eles representam máximos relativos, mínimos relativos ou pontos de inflexão.

2. Encontre os valores estacionários das seguintes funções:

(a) $y = (x - 1)^3 + 16$

(c) $y = (3 - x)^6 + 7$

(b) $y = (x - 2)^4$

(d) $y = (5 - 2x)^4 + 8$

Use o teste da N -ésima derivada para determinar a exata natureza desses valores estacionários.

[†] Se $f(x)$ for uma função constante, por exemplo, então, obviamente, $f'(x) = f''(x) = \dots = 0$, de modo que nunca poderá ser encontrado nenhum valor de derivada diferente de zero. Contudo, esse é um caso trivial, visto que, de qualquer maneira, uma função constante não requer nenhum teste para extremo. Como um exemplo não trivial, considere a função

$$y = \begin{cases} e^{-1/x^2} & (\text{para } x \neq 0) \\ 0 & (\text{para } x = 0) \end{cases}$$

onde a função $y = e^{-1/x^2}$ é uma função exponencial, que ainda será apresentada (Capítulo 10). Por si, $y = e^{-1/x^2}$ é descontínua em $x = 0$, porque $x = 0$ não está no domínio (divisão por zero é indefinida). Contudo, visto que $\lim_{x \rightarrow 0} y = 0$, se acrescentarmos a estipula-

ção $y = 0$ para $x = 0$, podemos preencher a lacuna no domínio e, assim, obter uma função contínua. O gráfico dessa função mostra que ela atinge um mínimo em $x = 0$. Contudo, acontece que, em $x = 0$, todas as derivadas (até qualquer ordem) têm valores nulos. Assim, não podemos aplicar o teste da N -ésima derivada para confirmar o fato – que pode ser provado graficamente – de que a função tem um mínimo em $x = 0$. Para uma discussão mais profunda desse caso excepcional, veja Courant, R. *Differential and Integral Calculus* (tradução de E. J. McShane), Nova York: Interscience, vol. I, 2. ed. p. 196, 197 e 336, 1937.



CAPÍTULO 10

Funções exponenciais e logarítmicas

O teste da N -ésima derivada desenvolvido no Capítulo 9 nos capacita para a tarefa de localizar os valores extremos de qualquer função objetivo, contanto que ela envolva somente uma variável de escolha, possua derivadas até a ordem desejada e, mais cedo ou mais tarde, resulte em uma derivada de valor diferente de zero no valor crítico x_0 . Nos exemplos citados no Capítulo 9, entretanto, utilizamos somente funções polinomiais e racionais para as quais sabemos como obter as derivadas necessárias. Suponha que nossa função objetivo seja, por acaso, uma função *exponencial*, como

$$y = 8^{x-\sqrt{x}}$$

Então continuamos incapazes de aplicar o critério de derivadas porque ainda não aprendemos como diferenciar tal função. E é isso que faremos neste capítulo.

As funções exponenciais, bem como as funções logarítmicas, estreitamente relacionadas com elas, têm aplicações importantes em economia, especialmente no que se refere a problemas de crescimento, e na dinâmica da economia em geral. A aplicação particular relevante para esta parte do livro, contudo, envolve uma classe de problemas de otimização na qual a variável de escolha é o *tempo*. Por exemplo, um certo comerciante de vinhos pode ter um estoque cujo valor de mercado sabe-se que aumentará com o tempo de alguma maneira predeterminada. O problema é determinar qual a melhor ocasião para vender aquele estoque tendo como base a função valor do vinho, levando em conta o custo de juros envolvido em manter o capital monetário vinculado àquele estoque. As funções exponenciais podem entrar nesse problema de dois modos. Primeiro, o valor do vinho pode aumentar com o tempo segundo alguma *lei exponencial de crescimento*. Se for esse o caso, teríamos uma função valor do vinho exponencial. Em segundo lugar, quando considerarmos o custo de juros, a presença de juros compostos com certeza introduzirá uma função exponencial no quadro. Por isso, devemos estudar a natureza das funções exponenciais antes de podermos discutir esse tipo de problema de otimização.

Uma vez que nosso propósito primordial é tratar o tempo como uma variável de escolha, agora vamos passar para o símbolo t —em lugar de x —para indicar a variável independente na discussão subsequente. (Todavia, esse mesmo símbolo também pode perfeitamente representar outras variáveis que não o tempo.)

10.1 A natureza das funções exponenciais

O termo *expoente*, já apresentado quando estudamos funções polinomiais, significa um indicador (índice) da potência à qual uma variável deve ser elevada. Em expressões de potências como x^3

ou x^5 , os expoentes são *constantes*; mas não há razão por que não possamos ter também um expoente *variável*, tal como em 3^x ou 3^t , onde o número 3 é elevado a potências variáveis (vários valores de x ou t). Uma função cuja variável *independente* aparece no papel de um expoente é denominada uma *função exponencial*.

Função exponencial simples

Em sua versão simples, a função exponencial pode ser representada na forma

$$y = f(t) = b^t \quad (b > 1) \quad (10.1)$$

onde y e t são, respectivamente, as variáveis dependente e independente e b denota uma *base* fixa do expoente. O domínio dessa função é o conjunto de todos os números reais. Assim, diferentemente dos expoentes em uma função polinomial, o expoente variável t em (10.1) não é limitado a inteiros positivos – a menos que queiramos impor tal restrição.

Mas, por que a restrição $b > 1$? A explicação é a seguinte. Visto que o domínio da função em (10.1) consiste no conjunto de todos os números reais, é possível que t assumira um valor como $\frac{1}{2}$. Se for permitido que b seja negativo, o fato de b estar elevado à potência meio envolverá tomar a raiz quadrada de um número negativo. Embora essa operação não seja impossível, com certeza é preferível a saída mais fácil, que é restringir b a ser positivo. Uma vez adotada a restrição $b > 0$, contudo, poderíamos perfeitamente ir mais além e adotar a restrição $b > 1$: a restrição $b > 1$ é diferente de $b > 0$ somente na exclusão adicional dos casos em que (1) $0 < b < 1$ e (2) $b = 1$; mas, como veremos em breve, o primeiro caso pode ser incorporado pela restrição $b > 1$, ao passo que o segundo caso pode ser descartado imediatamente. Considere o primeiro caso. Se $b = \frac{1}{5}$, então temos

$$y = \left(\frac{1}{5}\right)^t = \frac{1}{5^t} = 5^{-t}$$

Isso mostra que uma função cuja base é fracionária pode ser facilmente reescrita como uma função cuja base é maior que 1. Quanto ao segundo caso, o fato de $b = 1$ nos dará a função $y = 1^t = 1$, de modo que a função exponencial, na verdade, degenera para uma função constante; portanto, pode ser desqualificada como um membro da família exponencial.

Forma gráfica

O gráfico da função exponencial em (10.1) toma a forma geral da curva na Figura 10.1. A curva desenhada é baseada no valor $b = 2$; mas mesmo para outros valores de b prevalecerá a mesma configuração geral.

Podemos notar várias características importantes desse tipo de curva exponencial. Em primeiro lugar, ela é contínua e suave em toda a sua extensão; assim, a função deve ser continuamente diferenciável em toda a sua extensão. A propósito, ela é continuamente diferenciável qualquer número de vezes. Em segundo lugar, ela é estritamente crescente e, na verdade, y cresce a uma taxa crescente em toda a extensão da curva. Por conseqüência, as derivadas primeira e segunda da função $y = b^t$ devem ser positivas – um fato que devemos poder confirmar depois que desenvolvermos as fórmulas de diferenciação relevantes. Em terceiro lugar, notamos que, mesmo que o domínio da função contenha números negativos, bem como positivos, a faixa da função está limitada ao intervalo aberto $(0, \infty)$. Isto é, a variável dependente y é invariavelmente positiva, não importa qual o sinal da variável independente t .

A monotonicidade estrita da função exponencial tem, no mínimo, duas implicações interessantes e significativas. Em primeiro lugar, podemos inferir que a função exponencial deve ter uma função inversa que é, em si, estritamente monotônica. Veremos que essa função inversa é uma função *logarítmica*. Em segundo lugar, uma vez que monotonicidade estrita significa que há um único valor de t para um dado valor de y e, visto que a faixa da função exponencial é o intervalo $(0, \infty)$, deduz-se que podemos expressar *qualquer número positivo* como uma potência única

de base $b > 1$. Isso pode ser visto na Figura 10.1, onde a curva de $y = 2^t$ abrange todos os valores positivos de y em sua faixa; por conseguinte, qualquer valor positivo de y poderá ser expresso como uma potência única do número 2. Na verdade, mesmo que a base seja mudada para algum outro número real maior do 1, a mesma faixa é válida, de modo que é possível expressar qualquer número positivo y como uma potência de qualquer base $b > 1$.

Função exponencial generalizada

Este último ponto merece ser examinado mais de perto. Se um y positivo pode ser, de fato, expresso como potências de várias bases alternativas, então deve existir um procedimento geral de *conversão de base*. No caso da função $y = 9^t$, por exemplo, podemos transformá-la, de pronto, em $y = (3^2)^t = 3^{2t}$, convertendo, desse modo, a base de 9 para 3, contanto que o expoente seja devidamente alterado de t para $2t$. Essa mudança no expoente, necessária para a conversão de base, não cria nenhum novo tipo de função, pois, se fizermos que $w = 2t$, então $y = 3^{2t} = 3^w$ ainda está na forma de (10.1). Contudo, do ponto de vista da base 3, o expoente agora é $2t$ em vez de t . Qual é o efeito da adição de um coeficiente numérico (aqui, 2) ao expoente t ?

A resposta é encontrada na Figura 10.2a, onde estão desenhadas duas curvas – uma para a função $y = f(t) = b^t$ e outra para a função $y = g(t) = b^{2t}$. Visto que o expoente desta última é exatamente duas vezes o da primeira e, uma vez que é adotada uma base idêntica para as duas funções, a atribuição de um valor arbitrário $t = t_0$ na função g e de $t = 2t_0$ na função f deve resultar no mesmo valor:

$$f(2t_0) = g(t_0) = b^{2t_0} = y_0$$

Assim, a distância y_0J será metade de y_0K . Por raciocínio similar, para qualquer valor de y , a função g deve estar exatamente eqüidistante da função f e do eixo vertical. Podemos concluir, portanto, que *dobrar* o expoente tem o efeito de comprimir a curva exponencial *até uma distância eqüidistante* em relação ao eixo y , enquanto *dividir* o expoente *ao meio* afastará a curva do eixo y por *duas* vezes a distância horizontal.

É de interesse que ambas as funções compartilhem a mesma interseção vertical

$$f(0) = g(0) = b^0 = 1$$

A mudança do expoente t para $2t$, ou para qualquer outro múltiplo de t , não afetará a interseção vertical. Em termos de *compressão*, isso acontece porque comprimir uma distância horizontal zero ainda resultará em uma distância zero.

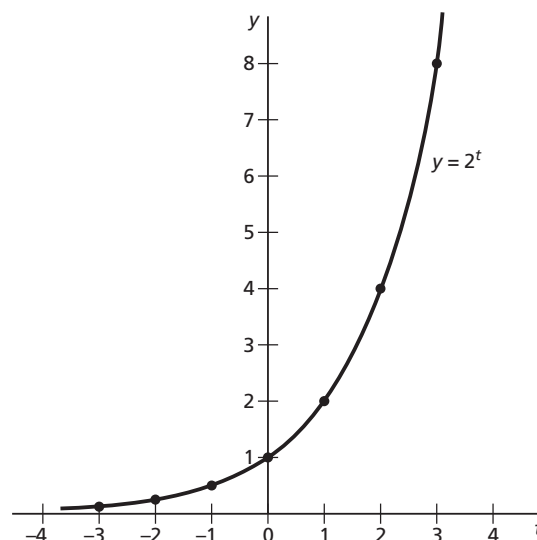


FIGURA 10.1

Mudar o expoente é uma maneira de modificar – e generalizar – a função exponencial de (10.1); uma outra maneira é anexar um coeficiente a b^t , tal como $2b^t$. [Atenção: $2b^t \neq (2b)^t$.] O efeito de tal coeficiente também é comprimir ou estender a curva, exceto que, dessa vez, a direção é vertical. Na Figura 10.2b, a curva mais alta representa $y = 2b^t$, e a mais baixa é $y = b^t$. Para todo valor de t , a primeira deve ser, obviamente, duas vezes mais alta, porque tem um valor de y duas vezes maior que a última. Assim, temos $t_0 J' = J' K'$. Note que a interseção vertical também muda no presente caso. Podemos concluir que *dobrar* o coeficiente (neste caso, de 1 para 2) serve para afastar a curva do eixo horizontal por *duas* vezes a distância vertical, ao passo que *dividir* o coeficiente *ao meio* comprimirá a curva para uma distância *equidistante* em direção ao eixo t .

Agora que conhecemos as duas modificações que acabamos de discutir, a função exponencial $y = b^t$ pode ser generalizada para a forma

$$y = ab^{ct} \quad (10.2)$$

onde a e c são agentes “de compressão” ou “de extensão”. Quando se atribuem valores diversos a esses agentes, eles alterarão a posição da curva exponencial, gerando, assim, toda uma família de curvas (funções) exponenciais. Se a e c forem positivos, prevalecerá a configuração geral mostrada na Figura 10.2; se a ou c ou ambos forem *negativos*, então ocorrerão modificações fundamentais na configuração da curva (veja Exercício 10.1-5).

Uma base preferida

O que sugeriu a discussão da mudança de expoente de t para ct foi a questão da conversão de base. Mas, dada a viabilidade da conversão da base, por que alguém desejaria fazê-la, afinal? Uma resposta é que algumas bases são mais convenientes que outras no que diz respeito a manipulações matemáticas.

O curioso é que, em cálculo, a base preferida é um número irracional denotado pelo símbolo e

$$e = 2,71828...$$

Quando essa base e é usada em uma função exponencial, ela é denominada uma *função exponencial natural*, e alguns exemplos dela são:

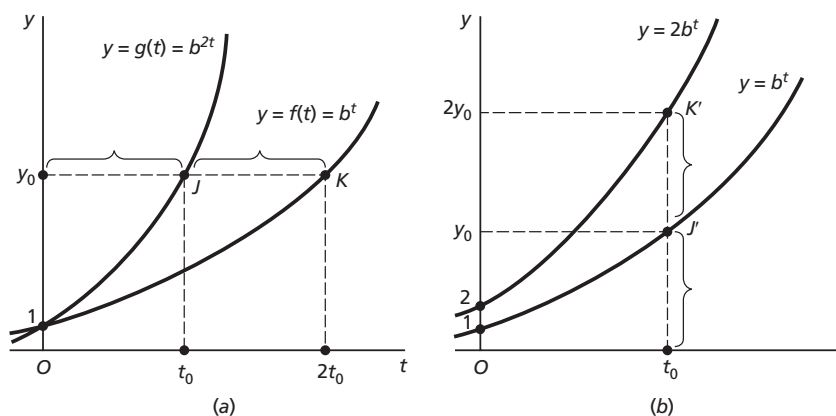
$$y = e^t \quad y = e^{3t} \quad y = Ae^{rt}$$

Essas funções ilustrativas também podem ser expressas pelas notações alternativas

$$y = \exp(t) \quad y = \exp(3t) \quad y = A \exp(rt)$$

onde a abreviação $\exp$ (para exponencial) indica que e deve ter como seu expoente a expressão entre parênteses.

FIGURA 10.2



A escolha de um número tão quanto como $e = 2,71828\dots$ como base preferida sem dúvida parece estranha. Mas há uma excelente razão para essa escolha, pois a função e^t possui a notável propriedade de ser sua própria derivada! Isto é,

$$\frac{d}{dt} e^t = e^t$$

um fato que reduz o trabalho de diferenciação a absolutamente nada. Além disso, tendo à mão essa regra de diferenciação – que será provada na Seção 10.5 –, também será fácil achar a derivada de uma função exponencial natural mais complicada, tal como $y = Ae^{rt}$. Para fazer isso, primeiro seja $w = rt$, de modo que a função se torna

$$y = Ae^w \quad \text{onde } w = rt, \text{ e } A, r \text{ são constantes}$$

Então, pela regra da cadeia, podemos escrever

$$\frac{dy}{dt} = \frac{dy}{dw} \frac{dw}{dt} = Ae^w(r) = r Ae^{rt}$$

Isto é,

$$\frac{d}{dt} Ae^{rt} = r Ae^{rt} \quad (10.3)$$

Assim, a conveniência matemática da base e deve ter ficado muito clara.

EXERCÍCIO 10.1

- Represente, em um único diagrama, os gráficos das funções exponenciais $y = 3^t$ e $y = 3^{2t}$.
 - Os dois gráficos apresentam a mesma relação geral entre posições mostrada na Figura 10.2a?
 - Essas curvas compartilham a mesma interseção y ? Por quê?
 - Esboce o gráfico da função $y = 3^{3t}$ no mesmo diagrama.
- Represente, em um único diagrama, os gráficos das funções exponenciais $y = 4^t$ e $y = 3(4^t)$.
 - Os dois gráficos apresentam a mesma relação geral entre posições sugerida na Figura 10.2b?
 - As duas curvas têm a mesma interseção y ? Por quê?
 - Esboce o gráfico da função $y = \frac{3}{2}(4^t)$ no mesmo diagrama.
- Admitindo como certo que e^t é sua própria derivada, use a regra da cadeia para achar dy/dt para as seguintes funções:
 - $y = e^{5t}$
 - $y = 4e^{3t}$
 - $y = 6e^{-2t}$
- Em vista de nossa discussão sobre (10.1), você espera que a função $y = e^t$ seja estritamente crescente a uma taxa crescente? Verifique sua resposta determinando os sinais das derivadas primeira e segunda dessa função. Ao fazer isso, lembre-se que o domínio dessa função é o conjunto de todos os números reais, isto é, o intervalo $(-\infty, \infty)$.
- Em (10.2), se forem atribuídos valores negativos a a e c , a forma geral das curvas na Figura 10.2 não mais prevalecerão. Examine a mudança na configuração da curva comparando (a) o caso de $a = -1$ com o caso de $a = 1$; e (b) o caso de $c = -1$ com o caso de $c = 1$.

10.2 Funções exponenciais naturais e o problema do crescimento

As questões pertinentes que ainda não foram respondidas são: como é definido o número e ? Esse número tem algum significado econômico além de sua importância matemática como uma base conveniente? E de que modo as funções exponenciais naturais se aplicam à análise econômica?

O número e

Vamos considerar a seguinte função:

$$f(m) = \left(1 + \frac{1}{m}\right)^m \quad (10.4)$$

Se forem atribuídos valores cada vez maiores a m , então $f(m)$ também assumirá valores maiores; especificamente, constatamos que

$$\begin{aligned} f(1) &= \left(1 + \frac{1}{1}\right)^1 = 2 \\ f(2) &= \left(1 + \frac{1}{2}\right)^2 = 2,25 \\ f(3) &= \left(1 + \frac{1}{3}\right)^3 = 2,37037... \\ f(4) &= \left(1 + \frac{1}{4}\right)^4 = 2,44141... \\ &\vdots \end{aligned}$$

Além disso, se m aumentar indefinidamente, então $f(m)$ convergirá para o número 2,71828... $\equiv e$; assim, e pode ser definido como o limite de (10.4) quando $m \rightarrow \infty$:

$$e \equiv \lim_{m \rightarrow \infty} f(m) = \lim_{m \rightarrow \infty} \left(1 + \frac{1}{m}\right)^m \quad (10.5)$$

O valor aproximado de e é 2,71828, o que pode ser verificado pelo cálculo da série de Maclaurin da função $\phi(x) = e^x$ – sendo que x é usado aqui para facilitar a aplicação direta da fórmula de expansão (9.14). Tal série nos dará uma aproximação polinomial de e^x e, assim, o valor de e ($= e^1$) pode ser aproximado estabelecendo $x = 1$ naquele polinômio. Se o resto R_n aproximar-se de zero à medida que o número de termos da série aumentar indefinidamente, isto é, se a série for convergente para $\phi(x)$, então podemos, de fato, aproximar o valor e com qualquer grau de exatidão desejado fazendo com que o número de termos incluídos seja suficientemente grande.

Para essa finalidade, precisamos ter derivadas de várias ordens para a função. Aceitando o fato de que a derivada primeira de e^x é o próprio e^x , podemos ver que a derivada de $\phi(x)$ é simplesmente e^x e, de modo semelhante, que as derivadas segunda, terceira ou de quaisquer ordens mais altas também devem ser e^x . Daí, quando avaliamos todas as derivadas no ponto de expansão ($x_0 = 0$), temos o gratificante e claro resultado

$$\phi'(0) = \phi''(0) = \dots = \phi^{(n)}(0) = e^0 = 1$$

Por conseqüência, estabelecendo $x_0 = 0$ em (9.14), a série de Maclaurin de e^x é

$$\begin{aligned} e^x = \phi(x) &= \phi(0) + \phi'(0)x + \frac{\phi''(0)}{2!}x^2 + \frac{\phi'''(0)}{3!}x^3 + \dots + \frac{\phi^{(n)}(0)}{n!}x^n + R_n \\ &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n + R_n \end{aligned}$$

De acordo com (9.15), o resto, R_n , pode ser escrito como

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!} x^{n+1} = \frac{e^p}{(n+1)!} x^{n+1} \quad [\phi^{(n+1)}(x) = e^x; \therefore \phi^{(n+1)}(p) = e^p]$$

Considerando que o valor da expressão fatorial $(n+1)!$ aumenta mais rapidamente que a expressão de potência x^{n+1} (para um x finito) à medida que n aumenta, deduz-se que $R_n \rightarrow 0$ quando $n \rightarrow \infty$. Assim, a série de Maclaurin converge e, por consequência, o valor de e^x pode ser expresso como uma série convergente infinita, como se segue:

$$e^x = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \frac{1}{4!} x^4 + \frac{1}{5!} x^5 + \dots \quad (10.6)$$

Como um caso especial, para $x = 1$, constatamos que

$$\begin{aligned} e &= 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \dots \\ &= 2 + 0,5 + 0,1666667 + 0,0416667 + 0,0083333 + 0,0013889 \\ &\quad + 0,0001984 + 0,0000248 + 0,0000028 + 0,0000003 + \dots \\ &= 2,7182819 \end{aligned}$$

Assim, se quisermos um número exato até a quinta casa decimal, podemos escrever $e = 2,71828$. Note que não precisamos nos preocupar com os termos subseqüentes na série infinita, porque suas magnitudes serão desprezíveis se estivermos preocupados somente com cinco casas decimais.

Uma interpretação econômica de e

Em termos matemáticos, o número e é o limite da expressão em (10.5). Mas ele também tem algum significado econômico? A resposta é que ele pode ser interpretado como o resultado de um modo especial de composição de juros.

Suponha que, começando com um principal (ou capital) de \$1, encontremos um banqueiro hipotético que nos ofereça a taxa de juros fora do comum de 100% ao ano (\$1 de juros por ano). Se o juro for composto uma vez por ano, o valor de nosso ativo ao final de um ano será \$2; denotaremos esse valor por $V(1)$, onde o número entre parênteses indica a freqüência de composição em 1 ano:

$$\begin{aligned} V(1) &= \text{principal inicial} (1 + \text{taxa de juros}) \\ &= 1(1 + 100\%) = (1 + \frac{1}{1})^1 = 2 \end{aligned}$$

Se for composto por semestre, entretanto, o juro resultante corresponderá a 50% (metade de 100%) do principal ao final de seis meses. Portanto, o principal será \$1,50 durante o segundo período de seis meses, no qual o juro será calculado a 50% de \$1,50. Assim, o valor de nosso ativo ao final do ano será $1,50(1 + 50\%)$; isto é,

$$V(2) = (1 + 50\%)(1 + 50\%) = (1 + \frac{1}{2})^2$$

Por raciocínio análogo, podemos escrever $V(3) = (1 + \frac{1}{3})^3$, $V(4) = (1 + \frac{1}{4})^4$ etc.; ou, de modo geral,

$$V(m) = \left(1 + \frac{1}{m}\right)^m \quad (10.7)$$

onde m representa a freqüência de composição ao longo de 1 ano.

No caso limite, quando o juro é composto *continuamente* durante o ano, isto é, quando m se torna infinito, o valor do ativo crescerá como uma “bola de neve” e, ao final de 1 ano, será

$$\lim_{m \rightarrow \infty} V(m) = \lim_{m \rightarrow \infty} \left(1 + \frac{1}{m}\right)^m = e \text{ (dólares)} \quad [\text{por (10.5)}]$$

Assim, o numero $e = 2,71828$ pode ser interpretado como o valor que um principal de \$1 alcançará ao final de 1 ano se o juro à taxa de 100% ao ano for composto continuamente.

Note que a taxa de juros de 100% é somente uma *taxa nominal de juro*, pois, se \$1 tornar-se \$ e = \$2,718 após 1 ano, a *taxa de juro efetiva* será, neste caso, aproximadamente 172% ao ano.

Juro composto e a função Ae^{rt}

O processo contínuo da composição de juros que acabamos de discutir pode ser generalizado em três direções, para levar em conta: (1) mais anos de composição; (2) outro principal que não \$1; e (3) uma taxa nominal de juro diferente de 100%.

Se um principal de \$1 tornar-se \$ e após um ano de juro composto contínuo e se deixarmos \$ e ser o novo principal no segundo ano (durante o qual cada dólar crescerá novamente para \$ e), o valor de nosso ativo ao final de 2 anos será, obviamente, \$ $e(e) = \$e^2$. Pelo mesmo critério, o valor será \$ e^3 ao final de 3 anos ou, em termos mais gerais, será \$ e^t após t anos.

A seguir, vamos mudar o principal de \$1 para uma quantidade não especificada, \$ A . Essa mudança é fácil de resolver: se \$1 crescerá até \$ e^t após t anos de composição contínua à taxa nominal de 100% ao ano, é óbvio que \$ A crescerá até \$ Ae^t .

E se a taxa nominal de juro for diferente de 100%, por exemplo, $r = 0,05 (= 5\%)$? O efeito dessa mudança de taxa é alterar a expressão \$ Ae^t para \$ Ae^{rt} , como podemos verificar a seguir. Com um principal inicial de \$ A para ser investido por t anos a uma taxa nominal de juro r , a fórmula de juro composto (10.7) deve ser modificada para a forma

$$V(m) = A \left(1 + \frac{r}{m}\right)^{mt} \quad (10.8)$$

A inserção do coeficiente A reflete a mudança do principal do nível anterior de \$1. A expressão quociente r/m significa que, em cada um dos m períodos de composição em um ano, somente $1/m$ da taxa nominal r será realmente aplicável. Por fim, o expoente mt nos diz que, visto que o juro tem de ser composto m vezes por ano, haverá um total de mt composições em t anos.

A fórmula (10.8) pode ser transformada para uma forma alternativa

$$\begin{aligned} V(m) &= A \left[\left(1 + \frac{r}{m}\right)^{m/t} \right]^{rt} \\ &= A \left[\left(1 + \frac{1}{w}\right)^w \right]^{rt} \quad \text{onde } w \equiv \frac{m}{r} \end{aligned} \quad (10.8')$$

Como a frequência de composição m aumentou, a nova variável w também deve crescer; assim, quando $m \rightarrow \infty$, temos $w \rightarrow \infty$, e a expressão entre colchetes em (10.8'), em virtude de (10.5), tende ao número e . Por consequência, constatamos que o valor do ativo no processo generalizado de composição contínua de juro é

$$V \equiv \lim_{m \rightarrow \infty} V(m) = Ae^{rt} \quad (10.8'')$$

como previsto.

Note que, em (10.8), t é uma variável *discreta* (no sentido oposto ao de *contínua*): ela só pode assumir valores que sejam múltiplos integrais de $1/m$. Por exemplo, se $m = 4$ (composto trimestralmente), então t pode assumir somente os valores de $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ etc., indicando que $V(m)$ assumirá um novo valor apenas ao final de cada novo trimestre. Contudo, quando $m \rightarrow \infty$, como em (10.8''), $1/m$ torna-se infinitesimal e, de acordo com isso, a variável t se tornará contínua. Nesse caso, é justo falar de frações de um ano e deixar t ser, digamos, 1,2 ou 2,35.

O resultado final é que as expressões e , e^t , Ae^t e Ae^{rt} podem ser interpretadas do ponto de vista da economia em conexão com o juro composto contínuo, como mostra a Tabela 10.1.

Taxa instantânea de crescimento

Devemos deixar claro, entretanto, que o juro composto é uma interpretação ilustrativa, mas não exclusiva, da função exponencial natural Ae^{rt} . O juro composto é um mero exemplo do processo geral de *crescimento exponencial* (aqui, o crescimento de uma soma de dinheiro (capital) ao longo do tempo), e podemos aplicar a função igualmente bem ao crescimento da população, da riqueza ou do capital real.

Aplicado a algum contexto que não seja o do juro composto, o coeficiente r em Ae^{rt} não mais designa a taxa nominal de juro. Qual significado econômico ele assume? A resposta é que r pode ser reinterpretado como a *taxa instantânea de crescimento* da função Ae^{rt} . (Na verdade, é por isso que adotamos o símbolo r , de “rate”, para taxa de crescimento). Dada a função $V = Ae^{rt}$, que dá o valor de V em cada ponto de tempo t , a taxa de variação de V deve ser encontrada na derivada

$$\frac{dV}{dt} = r Ae^{rt} = rV \quad [\text{veja (10.3)}]$$

Mas a *taxa de crescimento* de V é simplesmente a *taxa de variação em V* expressa em termos relativos (percentuais), isto é, expressa como uma razão relativa ao valor de V em si. Assim, para qualquer ponto de tempo dado, temos

$$\text{Taxa de crescimento de } V \equiv \frac{dV/dt}{V} = \frac{rV}{V} = r \quad (10.9)$$

como já foi dito.

Há diversas observações que devem ser feitas sobre essa taxa de crescimento. Mas, em primeiro lugar, vamos esclarecer um ponto fundamental referente ao conceito de tempo, a saber, a distinção entre um *ponto* de tempo e um *período* de tempo. A variável V (que denota uma soma em dinheiro, ou o tamanho da população etc.) é um conceito de *estoque*, que se preocupa com a pergunta: quanto dela *existe* em um dado momento? Como tal, V está relacionada ao conceito de *ponto* de tempo; a cada ponto de tempo, V assume um valor único. A variação em V , por outro lado, representa um *fluxo*, que envolve a pergunta: quanto dela *acontece* durante um dado período de tempo? Portanto, uma variação em V , pelo mesmo critério, a taxa de variação de V devem se referir a algum período de tempo especificado, digamos, por ano.

Isso posto e entendido, vamos voltar a (10.9) para alguns comentários:

1. A taxa de crescimento definida em (10.9) é uma taxa *instantânea* de crescimento. Visto que a derivada $dV/dt = r Ae^{rt}$ assume um valor diferente em um ponto diferente de t , o que

Principal, \$	Taxa nominal de juro	Anos de juro composto contínuo	Valor do ativo ao final do processo de juro composto, \$
1	100% (= 1)	1	e
1	100%	t	e^t
A	100%	t	Ae^t
A	r	t	Ae^{rt}

TABELA 10.1
Juro Composto Contínuo

- também acontece com $V = Ae^{rt}$, a razão entre elas também deve se referir a um ponto (ou *instante*) específico de t . Nesse sentido, a taxa de crescimento é instantânea.
2. No presente caso, entretanto, a taxa instantânea de crescimento por acaso é uma constante r e, assim, a taxa de crescimento permanece uniforme em todos os pontos de tempo. É claro que isso pode não ser verdade para todas as situações de crescimento que encontramos na prática.
 3. Embora a taxa de crescimento r seja medida em um ponto de tempo particular, sua grandeza tem a conotação de tantos por cento *por unidade de tempo*, digamos, por ano (se t for medido em unidades de anos). O crescimento, por sua própria natureza, só pode ocorrer em um intervalo de tempo. É por isso que uma única foto instantânea (que registra a situação em um instante) jamais poderia retratar, digamos, o crescimento de uma criança, ao passo que duas fotos instantâneas tiradas em épocas diferentes – digamos, no intervalo de um ano – conseguem fazer isso. Portanto, dizer que V tem uma taxa de crescimento r no instante $t = t_0$ significa realmente que, se for permitido que a taxa de variação $dV/dt (= rV)$ prevalecente em $t = t_0$ continue sem ser perturbada por toda uma unidade de tempo (1 ano), então V terá crescido de uma quantidade rV no final do ano.
 4. Para a função exponencial $V = Ae^{rt}$, a *taxa percentual de crescimento* é constante em todos os pontos t , mas a *quantidade absoluta* de incremento em V aumenta à medida que o tempo passa porque a taxa percentual será calculada sobre bases cada vez maiores.

Ao interpretar r como a taxa instantânea de crescimento, fica claro que, daqui em diante, pouco esforço será exigido para achar a taxa em crescimento de uma função exponencial natural na forma $y = Ae^{rt}$, contanto que r seja uma constante. Dada a função $y = 75e^{0,02t}$, por exemplo, podemos imediatamente deduzir que a taxa de crescimento de y é 0,02 ou 2% por período.

Crescimento contínuo versus crescimento discreto

A discussão precedente, embora interessante do ponto de vista analítico, ainda está aberta à controvérsia no que se refere à relevância econômica porque, na verdade, o crescimento nem sempre ocorre *de modo contínuo* – nem mesmo o juro composto. Porém, por sorte, mesmo para casos de crescimento *discreto*, onde ocorrem mudanças apenas uma vez por período e não de instante em instante, a utilização da função de crescimento exponencial contínuo pode ser justificável.

Uma razão é que, em casos em que a frequência de composição é relativamente alta, embora não infinita, o padrão contínuo de crescimento pode ser considerado como uma aproximação do verdadeiro padrão de crescimento. Porém, o mais importante é que podemos mostrar que um problema de crescimento discreto ou descontínuo sempre pode ser transformado em uma versão contínua equivalente.

Suponha que temos um padrão geométrico de crescimento (digamos, a composição *discreta* do juro) como mostra a seguinte seqüência:

$$A, A(1+i), A(1+i)^2, A(1+i)^3, \dots$$

onde a taxa efetiva de juro por período é denotada por i e onde o expoente da expressão $(1+i)$ denota o número de períodos abrangidos na composição. Se considerarmos $(1+i)$ como a base b em uma expressão exponencial, então a seqüência dada pode ser resumida pela função exponencial Ab^t – exceto que, por causa da natureza discreta do problema, t está restrito apenas a valores inteiros. Além disso, $b = 1+i$ é um número positivo (positivo mesmo que i seja uma taxa *negativa* de juro, digamos, $-0,04$), de modo que ela sempre pode ser expressa como uma potência de qualquer número real maior que 1, inclusive e . Isso significa que deve existir um número r tal que[†]

[†] O método para achar o número t , dado um valor específico de b , será discutido na Seção 10.4.

$$1 + i = b = e^r$$

Assim, podemos transformar Ab^t em uma função exponencial natural:

$$A(1 + i)^t = Ab^t = Ae^{rt}$$

Para qualquer valor de t dado – nesse contexto, valores inteiros de t – a função Ae^{rt} resultará, é claro, exatamente no mesmo valor que $A(1 + i)^t$, tal como $A(1 + i) = Ae^r$ e $A(1 + i)^2 = Ae^{2r}$. Por consequência, embora o caso considerado, $A(1 + i)^t$, seja *discreto*, ainda assim podemos trabalhar com a função exponencial natural *contínua* Ae^{rt} . Isso explica por que funções exponenciais naturais têm uma extensa aplicação em análise econômica, apesar do fato de que nem todos os padrões de crescimento são realmente contínuos.

Desconto e crescimento negativo

Agora vamos desviar um pouco nossa atenção do juro composto e passar para o conceito estreitamente relacionado de *desconto*. Em um problema de juro composto, procuramos calcular o *valor futuro* V (principal mais juros) dado um *valor presente* A (principal inicial). O problema do *desconto* é o oposto: achar o valor presente A de uma dada soma V que deverá estar disponível dentro de t anos.

Vamos examinar o caso discreto em primeiro lugar. Se o principal A crescer até o valor futuro de $A(1 + i)^t$ após t anos de juro composto anual à taxa i por ano, isto é, se

$$V = A(1 + i)^t$$

então, dividindo ambos os lados da equação pela expressão diferente de zero $(1 + i)^t$, podemos obter a fórmula do desconto:

$$A = \frac{V}{(1 + i)^t} = V(1 + i)^{-t} \quad (10.10)$$

que envolve um expoente negativo. É preciso entender que, nessa fórmula, os papéis de V e A foram invertidos: V agora é um dado, ao passo que A é a incógnita a ser calculada por i (a taxa de desconto) e t (o número de anos), bem como por V .

De modo similar, para o caso contínuo, se o principal A crescer até Ae^{rt} após t anos de juro composto à taxa r segundo a fórmula

$$V = Ae^{rt}$$

então podemos obter a fórmula de desconto contínuo correspondente simplesmente dividindo ambos os lados da última equação por e^{rt} :

$$A = \frac{V}{e^{rt}} = Ve^{-rt} \quad (10.11)$$

Aqui, mais uma vez, A (e não V) é a incógnita, a ser calculada a partir do valor futuro dado V , da taxa nominal de desconto r , e do número de anos t . A expressão e^{-rt} normalmente é denominada *fator de desconto*.

Tomando (10.11) como um função exponencial de crescimento, podemos ver imediatamente que $-r$ é a taxa instantânea de crescimento de A . Por ser negativa, essa taxa é, com efeito, uma *taxa de depreciação*. Exatamente como o juro composto é um exemplo do processo de crescimento, o desconto ilustra o crescimento *negativo*.

EXERCÍCIO 10.2

- Use a forma da série infinita de e^x em (10.6) para calcular o valor aproximado de:
 - e^2
 - $\sqrt{e}$ ($= e^{1/2}$)
 (Arredonde o cálculo de cada termo para três casas decimais e continue com a série até obter um termo 0,000.)
- Dada a função $\phi(x) = e^{2x}$:
 - Escreva a parte polinomial P_n de sua série de Maclaurin.
 - Escreva o resto R_n na forma de Lagrange. Determine se $R_n \rightarrow 0$ quando $n \rightarrow \infty$, isto é, se a série é convergente para $\phi(x)$.
 - Se for convergente, de modo que $\phi(x)$ possa ser expresso como uma série infinita, escreva essa série.
- Escreva uma expressão exponencial para o valor:
 - \$70, composto continuamente à taxa de juro de 4% por 3 anos
 - \$690, composto continuamente à taxa de juro de 5% por 2 anos
 (Essas taxas de juro são taxas nominais por ano.)
- Qual é a taxa instantânea de crescimento de y em cada uma das seguintes?
 - $y = e^{0,07t}$
 - $y = 15e^{0,03t}$
 - $y = Ae^{0,4t}$
 - $y = 0,03e^t$
- Mostre que as duas funções $y_1 = Ae^{rt}$ (juro composto) e $y_2 = Ae^{-rt}$ (desconto) são imagens especulares uma da outra em relação ao eixo y [veja Exercício 10.1-5, parte (b)].

10.3 Logaritmos

As funções exponenciais são estreitamente relacionadas com as *funções logarítmicas* (abreviadamente, *funções log*). Antes de podermos discutir as funções log, primeiro precisamos entender o significado do termo *logaritmo*.

O significado de logaritmo

Quando temos dois números, tais como 4 e 16, que podem ser relacionados um com o outro pela equação $4^2 = 16$, definimos o *expoente* 2 como o *logaritmo* de 16 na base 4, e escrevemos

$$\log_4 16 = 2$$

Esse exemplo deve deixar claro que o logaritmo nada mais é do que a *potência* à qual uma base (4) deve ser elevada para obter um número particular (16). Em geral, podemos afirmar que

$$y = b^t \quad \Leftrightarrow \quad t = \log_b y \quad (10.12)$$

o que indica que o log de y na base b (denotado por $\log_b y$) é a potência à qual a base b deve ser elevada para obter o valor y . Por essa razão, é correto, embora redundante, escrever

$$b^{\log_b y} = y$$

Dado y , o processo para achar seu logaritmo $\log_b y$ é denominado *tomar*, ou *achar*, o log de y na base b . O processo inverso, de achar y a partir de um valor conhecido de seu logaritmo $\log_b y$, é denominado *tomar*, ou *achar*, o *antilog* de $\log_b y$.

Quando discutimos funções exponenciais, enfatizamos que a função $y = b^t$ (com $b > 1$) é estritamente crescente. Isso significa que, para qualquer valor positivo de y , há um *único* expoente t (não necessariamente positivo) tal que $y = b^t$; além disso, quanto maior o valor de y , maior deve ser t , como pode ser visto na Figura 10.2. Traduzindo para logaritmos, a monotonicidade estrita da função exponencial implica que qualquer número positivo y deve possuir um *único* logaritmo

t em uma base $b > 1$ tal que, quanto maior o valor de y , maior seu logaritmo. Como mostram as Figuras 10.1 e 10.2, y é necessariamente positivo na função exponencial $y = b^t$; conseqüentemente, um número negativo ou zero não possui um logaritmo.

Log comum e log natural

A base do logaritmo, $b > 1$, não tem de ser restrita a qualquer número particular, mas, em aplicações propriamente ditas de log, há dois números que são amplamente escolhidos como bases – o número 10 e o número e . Quando a base é 10, o logaritmo é conhecido como *logaritmo comum*, simbolizado por $\log_{10}$ (ou, se o contexto for claro, simplesmente por $\log$). Se a base for o número e , por outro lado, o logaritmo é denominado *logaritmo natural* e é representado ou por $\log_e$ ou por $\ln$ (de log natural). Também podemos usar o símbolo $\log$ (sem o subscrito e) se não ficar ambíguo no contexto específico.

Logaritmos comuns, usados com freqüência em *cálculos*, são exemplificados a seguir:

$$\begin{aligned}\log_{10} 1000 &= 3 && [\text{porque } 10^3 = 1000] \\ \log_{10} 100 &= 2 && [\text{porque } 10^2 = 100] \\ \log_{10} 10 &= 1 && [\text{porque } 10^1 = 10] \\ \log_{10} 1 &= 0 && [\text{porque } 10^0 = 1] \\ \log_{10} 0,1 &= -1 && [\text{porque } 10^{-1} = 0,1] \\ \log_{10} 0,01 &= -2 && [\text{porque } 10^{-2} = 0,01]\end{aligned}$$

Observe a estreita relação entre o conjunto de números imediatamente à esquerda dos sinais de igualdade e o conjunto de números imediatamente à direita. Desses exemplos, deve ficar aparente que o logaritmo comum de um número entre 10 e 100 deve estar entre 1 e 2 e que o logaritmo comum de um número entre 1 e 10 deve ser uma fração positiva etc. Os logaritmos exatos podem ser obtidos com facilidade em uma tabela (ou tábua) de logaritmos comuns ou em calculadoras habilitadas para cálculos com $\log$.[†]

No trabalho *analítico*, entretanto, os logaritmos naturais demonstram ser muito mais convenientes que os logaritmos comuns. Visto que, pela definição de logaritmo, temos a relação

$$y = e^t \quad \Leftrightarrow \quad t = \log_e y \quad (\text{ou } t = \ln y) \quad (10.13)$$

é fácil ver que a conveniência analítica de e em funções exponenciais será automaticamente estendida para o reino dos logaritmos de base e .

Os exemplos seguintes servirão para ilustrar os logaritmos naturais:

$$\begin{aligned}\ln e^3 &= \log_e e^3 = 3 \\ \ln e^2 &= \log_e e^2 = 2 \\ \ln e^1 &= \log_e e^1 = 1 \\ \ln 1 &= \log_e e^0 = 0 \\ \ln \frac{1}{e} &= \log_e e^{-1} = -1\end{aligned}$$

Por esses exemplos, podemos perceber que o princípio geral é que, dada uma expressão e^k , onde k é qualquer número real, podemos ler automaticamente o expoente k como o log natural de e^k . Portanto, em geral, o resultado é que $\ln e^k = k$.^{††}

[†] Em sentido mais fundamental, o valor de um logaritmo, assim como o valor de e , pode ser calculado (ou aproximado) recorrendo-se a uma expansão de Maclaurin de uma função $\log$, de um modo semelhante ao esboçado em (10.6). Contudo, não vamos nos aventurar nessa direção.

^{††} Como recurso mnemônico, observe que, quando o símbolo $\ln$ (ou $\log_e$) estiver à esquerda da expressão e^k , o símbolo $\ln$ parece cancelar o símbolo, deixando k como a resposta.

O log comum e o log natural podem ser convertidos um no outro; isto é, a base de um logaritmo pode ser mudada, exatamente como uma expressão exponencial. Desenvolveremos um par de fórmulas de conversão depois de estudarmos as regras básicas de logaritmos.

Regras de logaritmos

Logaritmos têm as características de expoentes; por conseguinte, obedecem certas regras estreitamente relacionadas com as regras de expoentes apresentadas na Seção 2.5. Elas podem ser de grande auxílio, pois simplificam as operações matemáticas. As três primeiras regras são enunciadas somente em termos do log natural, mas também são válidas quando o símbolo $\ln$ for substituído por $\log_b$.

Regra I (log de um produto) $\ln(uv) = \ln u + \ln v \quad (u, v > 0)$

EXEMPLO 1

$$\ln(e^6 e^4) = \ln e^6 + \ln e^4 = 6 + 4 = 10$$

EXEMPLO 2

$$\ln(Ae^7) = \ln A + \ln e^7 = \ln A + 7$$

DEMONSTRAÇÃO Por definição, $\ln u$ é a potência à qual e deve ser elevado para obter o valor de u ; assim, $e^{\ln u} = u$.[†] De modo semelhante, temos $e^{\ln v} = v$ e $e^{\ln(uv)} = uv$. A última é uma expressão exponencial para uv . Contudo, pode-se obter uma outra expressão para uv pela multiplicação direta de u e v :

$$uv = e^{\ln u} e^{\ln v} = e^{\ln u + \ln v}$$

Assim, igualando as duas expressões para uv acima, encontramos

$$e^{\ln(uv)} = e^{\ln u + \ln v} \quad \text{e, portanto} \quad \ln(uv) = \ln u + \ln v$$

Regra II (log de um quociente) $\ln(u/v) = \ln u - \ln v \quad (u, v > 0)$

EXEMPLO 3

$$\ln(e^2/c) = \ln e^2 - \ln c = 2 - \ln c$$

EXEMPLO 4

$$\ln(e^2/e^5) = \ln e^2 - \ln e^5 = 2 - 5 = -3$$

A demonstração dessa regra é muito semelhante à da Regra I e, portanto, vamos deixá-la como exercício.

Regra III (log de uma potência) $\ln u^a = a \ln u \quad (u > 0)$

EXEMPLO 5

$$\ln e^{15} = 15 \ln e = 15$$

EXEMPLO 6

$$\ln A^3 = 3 \ln A$$

DEMONSTRAÇÃO Por definição, $e^{\ln u} = u$; e, de modo semelhante $e^{\ln u^a} = u^a$. Todavia, podemos formar uma outra expressão para u^a , como se segue:

$$u^a = (e^{\ln u})^a = e^{a \ln u}$$

Igualando os expoentes das duas expressões para u^a , obtemos o resultado desejado, $\ln u^a = a \ln u$.

[†] Note que, quando e for elevado à potência $\ln u$, o símbolo e e o símbolo $\ln$ mais uma vez parecem ser cancelados, deixando u como a resposta.

Essas três regras são dispositivos úteis para simplificar as operações matemáticas em certos tipos de problemas. A Regra I serve para converter, por meio de logaritmos, uma operação de multiplicação (uv) em uma operação de adição ($\ln u + \ln v$); a Regra II transforma uma divisão (u/v) em uma subtração ($\ln u - \ln v$); e a Regra III nos habilita a reduzir uma potência a uma constante multiplicativa. Além disso, essas regras podem ser usadas em conjunto. E também podem ser lidas de trás para a frente e aplicadas ao contrário.

EXEMPLO 7

$$\ln(uv^a) = \ln u + \ln v^a = \ln u + a \ln v$$

EXEMPLO 8

$$\ln u + a \ln v = \ln u + \ln v^a = \ln(uv^a) \quad [\text{Exemplo 7 ao contrário}]$$

Entretanto, aqui cabe um alerta: quando tivermos expressões *aditivas* de início, os logaritmos não podem nos ajudar em nada. Em particular, devemos lembrar que

$$\ln(u \pm v) \neq \ln u \pm \ln v$$

Agora vamos apresentar duas regras adicionais referentes à mudança de base de um logaritmo.

Regra IV (conversão de base de log) $\log_b u = (\log_b e)(\log_e u) \quad (u > 0)$

Essa regra, que lembra a regra da cadeia (observe a “cadeia” $b \nearrow^e \searrow_e u$), nos habilita a calcular o logaritmo $\log_e u$ (na base e) a partir do logaritmo $\log_b u$ (na base b), ou vice-versa.

DEMONSTRAÇÃO Seja $u = e^p$, de modo que $p = \log_e u$. Então, deduz-se que

$$\log_b u = \log_b e^p = p \log_b e = (\log_e u)(\log_b e)$$

A Regra IV pode ser generalizada, de pronto, para

$$\log_b u = (\log_b c)(\log_c u)$$

onde c é alguma outra base que não b .

Regra V (inversão de base de log) $\log_b e = \frac{1}{\log_e b}$

Essa regra, que é parecida com a regra de diferenciação de função inversa, nos habilita a obter o log de b na base e imediatamente após dado o log de e na base b , e vice-versa. (Essa regra também pode ser generalizada para a forma $\log_b c = 1/\log_c b$.)

DEMONSTRAÇÃO Como uma aplicação da Regra IV, seja $u = b$; então, temos

$$\log_b b = (\log_b e)(\log_e b)$$

Mas a expressão do lado esquerdo é $\log_b b = 1$; portanto, $\log_b e$ e $\log_e b$ devem ser recíprocos um do outro, como afirma a Regra V.

Das duas últimas regras é fácil derivar o seguinte par de fórmulas de conversão entre log comum e log natural:

$$\begin{aligned} \log_{10} N &= (\log_{10} e)(\log_e N) = 0,4343 \log_e N \\ \log_e N &= (\log_e 10)(\log_{10} N) = 2,3026 \log_{10} N \end{aligned} \quad (10.14)$$

para N , um número positivo real. O primeiro sinal de igualdade de cada fórmula é facilmente justificável pela Regra IV. Na primeira fórmula, o valor 0,4343 (o log comum de 2,71828) pode ser encontrado em uma tabela de logaritmos comuns ou em uma calculadora eletrônica; na segunda, o valor 2,3026 (o log natural de 10) é apenas o recíproco de 0,4343, calculado desse modo por causa da Regra V.

EXEMPLO 9 $\log_e 100 = 2,3026(\log_{10} 100) = 2,3026(2) = 4,6052$. Reciprocamente, temos $\log_{10} 100 = 0,4343$ ($\log_e 100) = 0,4343(4,6052) = 2$.

Uma aplicação

As regras de logaritmos precedentes nos habilitam a resolver com facilidade certas *equações exponenciais* simples (*funções* exponenciais igualadas a zero). Por exemplo, se quisermos achar o valor de x que satisfaz a equação

$$ab^x - c = 0 \quad (a, b, c > 0)$$

podemos, em primeiro lugar, tentar transformar essa equação exponencial, utilizando logaritmos, em uma equação *linear* e então resolvê-la como tal. Para essa finalidade, primeiro o termo c deve ser transferido para o lado direito:

$$ab^x = c$$

Isso porque não há nenhuma expressão logarítmica simples para a expressão aditiva ($ab^x - c$), mas existem expressões logarítmicas convenientes para o termo multiplicativo ab^x e para c individualmente. Assim, após transpor c e tomar o logaritmo (digamos, na base 10) de ambos os lados, temos

$$\log a + x \log b = \log c$$

que é uma equação linear com variável x , cuja solução é

$$x = \frac{\log c - \log a}{\log b}$$

EXERCÍCIO 10.3

- Quais são os valores dos seguintes logaritmos?

(a) $\log_{10} 10.000$	(c) $\log_3 81$
(b) $\log_{10} 0,0001$	(d) $\log_5 3.125$
- Calcule os seguintes:

(a) $\ln e^7$	(c) $\ln(1/e^3)$	(e) $(e^{\ln 3})!$
(b) $\log_e e^{-4}$	(d) $\log_e(1/e^2)$	(f) $\ln e^x - e^{\ln x}$
- Calcule os seguintes aplicando as regras de logaritmos:

(a) $\log_{10}(100)^{13}$	(c) $\ln(3/B)$	(e) $\ln ABe^{-4}$
(b) $\log_{10} \frac{1}{100}$	(d) $\ln Ae^2$	(f) $(\log_4 e)(\log_e 64)$
- Quais das expressões seguintes são válidas?

(a) $\ln u - 2 = \ln \frac{u}{e^2}$	(c) $\ln u + \ln v - \ln w = \ln \frac{uv}{w}$
(b) $3 + \ln v = \frac{e^3}{v}$	(d) $\ln 3 + \ln 5 = \ln 8$
- Prove que $\ln(u/v) = \ln u - \ln v$.

10.4 Funções logarítmicas

Quando uma variável é expressa como uma função do logaritmo de uma outra variável, a função é denominada uma *função logarítmica*. Já vimos duas versões desse tipo de função em (10.12) e (10.13), a saber,

$$t = \log_b y \quad \text{e} \quad t = \log_e y (= \ln y)$$

que são diferentes uma da outra somente no que se refere à base do logaritmo.

Funções log e funções exponenciais

Como já dissemos antes, as funções log são funções inversas de certas funções exponenciais. O exame das duas funções logarítmicas anteriores confirmará que elas são, de fato, as respectivas funções inversas das funções exponenciais

$$y = b^t \quad \text{e} \quad y = e^t$$

porque as funções log citadas são o resultado da inversão de papéis das variáveis dependente e independente das funções exponenciais correspondentes. Você deve perceber, é claro, que o símbolo t está sendo usado aqui como um símbolo geral, e não representa, necessariamente, *tempo*. Mesmo quando representa, sua presença como variável *dependente* não significa que o tempo é determinado por alguma variável y ; significa apenas que um dado valor de y é associado a um único ponto de tempo.

Sendo funções inversas de funções estritamente crescentes (exponenciais), as funções logarítmicas também devem ser estritamente crescentes, o que é consistente com o que declaramos antes, isto é, quanto maior o número, maior será seu logaritmo em qualquer base dada. Essa propriedade pode ser expressa simbolicamente em termos das duas proposições seguintes: para dois valores positivos de y (y_1 e y_2),

$$\begin{aligned} \ln y_1 = \ln y_2 & \Leftrightarrow y_1 = y_2 \\ \ln y_1 > \ln y_2 & \Leftrightarrow y_1 > y_2 \end{aligned} \quad (10.15)$$

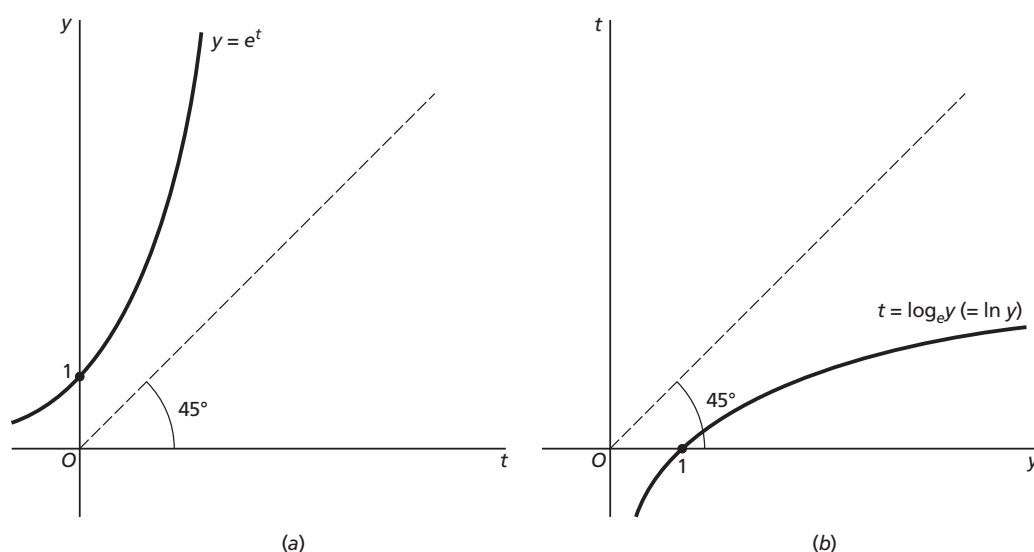
Essas proposições também são válidas, é claro, se substituirmos $\ln$ por $\log_b$.

A forma gráfica

A monotonicidade e outras propriedades gerais das funções logarítmicas podem ser observadas com clareza em seus gráficos. Dado o gráfico da função exponencial $y = e^t$, podemos obter o gráfico da função log correspondente traçando novamente o gráfico original, porém transpondo os dois eixos. O resultado dessa operação é ilustrado na Figura 10.3. Note que, se o gráfico da Figura 10.3b fosse sobreposto ao gráfico da Figura 10.3a, com o eixo y sobre o eixo t e o eixo t sobre o eixo y , as duas curvas coincidiriam exatamente. Por outro lado, do modo como aparecem na Figura 10.3 – com os eixos trocados – as duas curvas se apresentam como imagens especulares uma da outra (como devem ser os gráficos de qualquer par de funções inversas) em relação à reta a 45° que passa pela origem.

Essa relação de imagens especulares tem várias implicações valiosas. Uma é que, embora ambas sejam estritamente crescentes, a curva logarítmica mostra y crescendo a uma *taxa decrescente* (derivada segunda negativa), distintamente ao contrário da curva exponencial, que mostra y aumentando a uma taxa crescente. Um outro contraste interessante é que, enquanto a função exponencial tem uma *faixa* positiva, a função logarítmica tem, por sua vez, um *domínio* positivo. (Essa última restrição sobre o domínio da função logarítmica é, obviamente, apenas um outro modo de dizer que somente números positivos têm logaritmos). Uma terceira consequência da relação de imagens especulares é que, exatamente como $y = e^t$ tem uma interseção vertical em 1, a

FIGURA 10.3



função logarítmica $t = \log_e y$ deve cortar o eixo horizontal em $y = 1$, indicando que $\log_e 1 = 0$. Considerando que essa interseção horizontal não é afetada pela base do logaritmo – por exemplo, $\log_{10} 1 = 0$, também – podemos inferir, da forma geral da curva logarítmica na Figura 10.3b, que, para *qualquer* base,

$$\left. \begin{array}{l} 0 < y < 1 \\ y = 1 \\ y > 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \log y < 0 \\ \log y = 0 \\ \log y > 0 \end{array} \right. \quad (10.16)$$

À guisa de verificação, podemos examinar os dois conjuntos de exemplos de logaritmos comuns e naturais dados na Seção 10.3. Além do mais, podemos notar que

$$\log y \rightarrow \begin{cases} \infty \\ -\infty \end{cases} \quad \text{quando } y \rightarrow \begin{cases} \infty \\ 0^+ \end{cases} \quad (10.16')$$

A comparação gráfica entre a função logarítmica e a função exponencial na Figura 10.3 é baseada nas funções simples $y = e^t$ e $t = \ln y$. O mesmo resultado geral prevalecerá se compararmos a função exponencial generalizada $y = Ae^{rt}$ com sua função logarítmica correspondente. Mesmo com as constantes (positivas) A e r para *comprimir* ou *estender* a curva exponencial, ela será parecida com a forma geral da Figura 10.3a, exceto que sua interseção vertical será em $y = A$ em vez de em $y = 1$ (quando $t = 0$, temos $y = Ae^0 = A$). De acordo com isso, sua função inversa deve ter uma interseção *horizontal* em $y = A$. De modo geral, no que diz respeito à reta a 45° , a curva logarítmica correspondente será uma imagem especular da curva exponencial.

Se quisermos a expressão algébrica específica da inversa de $y = Ae^{rt}$, ela pode ser obtida tomando-se o log natural de ambos os lados dessa função exponencial [o que, de acordo com a primeira proposição em (10.15), não causará nenhuma perturbação na equação] e então resolvendo para t :

$$\ln y = \ln(Ae^{rt}) = \ln A + rt \ln e = \ln A + rt$$

Portanto,

$$t = \frac{\ln y - \ln A}{r} \quad (r \neq 0) \quad (10.17)$$

Esse resultado, uma função logarítmica, constitui a inversa da função exponencial $y = Ae^{rt}$. Como afirmamos antes, a função em (10.17) tem uma interseção horizontal em $y = A$, porque, quando $y = A$, temos $\ln y = \ln A$ e, portanto, $t = 0$.

Conversão de base

Na Seção 10.2, declaramos que a função exponencial $y = Ab^t$ sempre pode ser convertida para uma função exponencial *natural* $y = Ae^{rt}$. Agora já estamos prontos para obter uma fórmula de conversão. Contudo, em vez de Ab^t , vamos considerar a conversão da expressão mais geral Ab^{ct} para Ae^{rt} . Visto que a essência do problema é achar um r a partir de valores dados de b e c tais que

$$e^r = b^c$$

basta expressar r como uma função de b e c . Essa tarefa é fácil de executar tomando-se o log natural de ambos os lados da última equação:

$$\ln e^r = \ln b^c$$

O lado esquerdo pode ser lido imediatamente como igual a r , de modo que a função desejada (fórmula de conversão) surge como

$$r = \ln b^c = c \ln b \quad (10.18)$$

Isso indica que a função $y = Ab^{ct}$ sempre pode ser reescrita na forma de base natural, $y = Ae^{(c \ln b)t}$.

EXEMPLO 1

Converta $y = 2^t$ em uma função exponencial natural. Aqui, temos $A = 1$, $b = 2$ e $c = 1$. Portanto, $r = c \ln b = \ln 2$, e a função exponencial desejada é

$$y = Ae^{rt} = e^{(\ln 2)t}$$

Se quisermos, também podemos calcular o valor numérico de $\ln 2$ por meio da utilização de (10.14) e de uma tabela de logaritmos comuns, como se segue:

$$\ln 2 = 2,3026 \log_{10} 2 = 2,3026(0,3010) = 0,6931 \quad (10.19)$$

Então, podemos expressar o resultado anterior alternativamente como $y = e^{0,6931t}$.

EXEMPLO 2

Converta $y = 3(5)^{2t}$ em uma função exponencial natural. Neste exemplo, $A = 3$, $b = 5$ e $c = 2$, e a fórmula (10.18) nos dá $r = 2 \ln 5$. Portanto, a função desejada é

$$y = Ae^{rt} = 3e^{(2 \ln 5)t}$$

Mais uma vez, se quisermos, podemos calcular

$$2 \ln 5 = \ln 25 = 2,3026 \log_{10} 25 = 2,3026(1,3979) = 3,2188$$

de modo que o resultado anterior pode ser expresso alternativamente como $y = 3e^{3,2188t}$.

É claro que também é possível converter funções logarítmicas na forma $t = \log_b y$ em funções logarítmicas naturais equivalentes. Para fazer isso, basta aplicar a Regra IV de logaritmos, que pode ser expressa como

$$\log_b y = (\log_b e)(\log_e y)$$

A substituição direta desse resultado na função logarítmica dada nos dá, imediatamente, a função logarítmica natural desejada:

$$\begin{aligned}
 t &= \log_b y = (\log_b e)(\log_e y) \\
 &= \frac{1}{\log_e b} \log_e y \quad [\text{pela Regra V de logaritmos}] \\
 &= \frac{\ln y}{\ln b}
 \end{aligned}$$

Pelo mesmo procedimento, podemos transformar a função logarítmica mais geral $t = a \log_b(cy)$ na forma equivalente

$$t = a(\log_b e)(\log_e cy) = \frac{a}{\log_e b} \log_e(cy) = \frac{a}{\ln b} \ln(cy)$$

EXEMPLO 3 Converta a função $t = \log_2 y$ em uma forma logarítmica natural. Visto que, neste exemplo, temos $b = 2$ e $a = c = 1$, a função desejada é

$$t = \frac{1}{\ln 2} \ln y$$

Por (10.19), contudo, também podemos expressá-la como $t = (1/0,6931) \ln y$.

EXEMPLO 4 Converta a função $t = 7 \log_{10}(2y)$ em uma função logarítmica natural. Neste caso, os valores das constantes são $a = 7$, $b = 10$ e $c = 2$; por consequência, a função desejada é

$$t = \frac{7}{\ln 10} \ln(2y)$$

Mas, visto que $\ln 10 = 2,3026$, como indica por (10.14), essa função pode ser reescrita como $t = (7/2,3026) \ln(2y) = 3,0400 \ln(2y)$.

Na discussão precedente, adotamos a prática de expressar t como uma função de y quando a função é logarítmica. A única razão para fazer isso é nosso desejo de destacar a relação de função inversa entre as funções exponenciais e logarítmicas. Quando uma função logarítmica for estudada *por si só*, nos a escreveremos $y = \ln t$ (em vez de $t = \ln y$), como de costume. Naturalmente, no que diz respeito ao aspecto analítico da discussão, nada será afetado por essa troca de símbolos.

EXERCÍCIO 10.4

1. A forma da função inversa de $y = Ae^{rt}$ em (10.17) requer que r seja diferente de zero. Qual é o significado desse requisito quando o consideramos em relação à função exponencial original $y = Ae^{rt}$?
2. (a) Esboce um gráfico da função exponencial $y = Ae^{rt}$; indique o valor da interseção vertical.
(b) Então esboce o gráfico da função logarítmica $t = \frac{\ln y - \ln A}{r}$, e indique o valor da interseção horizontal.
3. Obtenha a função inversa de $y = ab^{ct}$.
4. Transforme as seguintes funções em suas formas exponenciais naturais:
(a) $y = 8^{3t}$ (c) $y = 5(5)^t$
(b) $y = 2(7)^{2t}$ (d) $y = 2(15)^{4t}$
5. Transforme as seguintes funções em suas formas logarítmicas naturais:
(a) $t = \log_7 y$ (c) $t = 3 \log_{15}(9y)$
(b) $t = \log_8(3y)$ (d) $t = 2 \log_{10} y$

6. Obtenha a taxa nominal de juro composto contínuo por ano (r) equivalente a uma taxa discreta de juro composto (i) de
 - (a) 5% ao ano, composto por ano.
 - (b) 5% ao ano, composto por semestre.
 - (c) 6% ao ano, composto por semestre.
 - (d) 6% ao ano, composto por trimestre.
7. (a) Ao descrever a Figura 10.3, o texto declara que, se as duas curvas forem sobrepostas uma à outra, elas mostram uma relação de imagem especular. Onde está localizado o “espelho”?
 - (b) Se traçarmos uma função $f(x)$ e sua negativa, $-f(x)$, no mesmo diagrama, as duas curvas também mostrarão uma relação de imagem especular? Se a resposta for positiva, onde está localizado o “espelho” neste caso?
 - (c) Se traçarmos os gráficos de Ae^{rt} e Ae^{-rt} no mesmo diagrama, as duas curvas serão imagens especulares uma da outra? Se forem, onde está localizado o “espelho”?

10.5 Derivadas de funções exponenciais e logarítmicas

Dissemos anteriormente que a função e^t é sua própria derivada. Na verdade, a função logarítmica natural, $\ln t$, também possui uma derivada bem conveniente, a saber, $d(\ln t)/dt = 1/t$. Esse fato reforça nossa preferência pela base e . Agora vamos provar a validade dessas duas fórmulas de derivadas e então deduziremos as fórmulas de derivadas para certas variantes das expressões exponencial e logarítmica e^t e $\ln t$.

Regra da função logarítmica

A derivada da função logarítmica $y = \ln t$ é

$$\frac{d}{dt} \ln t = \frac{1}{t}$$

Para provar isso, lembremos que, por definição, a derivada de $y = \psi(t) = \ln t$ tem o seguinte valor em $t = N$ (admitindo $t \rightarrow N^+$):

$$\begin{aligned} \psi'(N) &= \lim_{t \rightarrow N^+} \frac{\psi(t) - \psi(N)}{t - N} = \lim_{t \rightarrow N^+} \frac{\ln t - \ln N}{t - N} \\ &= \lim_{t \rightarrow N^+} \frac{\ln(t/N)}{t - N} \quad [\text{pela Regra II de logaritmos}] \end{aligned}$$

Agora vamos introduzir um símbolo abreviado $m \equiv \frac{N}{t - N}$. Então podemos escrever $\frac{1}{t - N} = \frac{m}{N}$, e também $\frac{t}{N} = 1 + \frac{t - N}{N} = 1 + \frac{1}{m}$. Assim, a expressão à direita do sinal de limite na equação anterior pode ser convertida para a forma

$$\frac{1}{t - N} \ln \frac{t}{N} = \frac{m}{N} \ln \left(1 + \frac{1}{m} \right) = \frac{1}{N} \ln \left(1 + \frac{1}{m} \right)^m \quad [\text{pela Regra III de logaritmos}]$$

Note que, quando $t \rightarrow N^+$, m tende ao infinito. Assim, para achar o valor desejado da derivada, podemos tomar o limite da última expressão na equação precedente, quando $m \rightarrow \infty$:

$$\psi'(N) = \lim_{m \rightarrow \infty} \frac{1}{N} \ln \left(1 + \frac{1}{m} \right)^m = \frac{1}{N} \ln e = \frac{1}{N} \quad [\text{por (10.5)}]$$

Entretanto, visto que N pode ser qualquer número para o qual um logaritmo é definido, podemos generalizar esse resultado e escrever $\psi'(t) = d(\ln t)/dt = 1/t$. Isso prova a regra da função logarítmica para $t \rightarrow N^+$.

O caso de $t \rightarrow N^-$ necessita de algumas modificações, mas a essência da demonstração é semelhante. Agora, a derivada de $y = \ln t$ tem o valor

$$\begin{aligned} \psi'(N) &= \lim_{t \rightarrow N^+} \frac{\psi(t) - \psi(N)}{t - N} = \lim_{t \rightarrow N^+} \frac{\psi(N) - \psi(t)}{N - t} \\ &= \lim_{t \rightarrow N^-} \frac{\ln N - \ln t}{N - t} = \lim_{t \rightarrow N^-} \frac{\ln(N/t)}{N - t} \end{aligned}$$

Seja $\mu = t/(N-t)$; então, $1/(N-t) = \mu/t$, e $N/t = 1 + (N-t)/t = 1 + 1/\mu$. Essas equações nos habilitam a reescrever a expressão à direita do último sinal de limite na equação precedente para $\psi'(N)$ como

$$\frac{1}{N-t} \ln \frac{N}{t} = \frac{\mu}{t} \ln \left(1 + \frac{1}{\mu} \right) = \frac{1}{t} \ln \left(1 + \frac{1}{\mu} \right)^\mu$$

Quando $t \rightarrow N^-$, $\mu \rightarrow \infty$. Assim, o valor da derivada desejada é

$$\psi'(N) = \lim_{t \rightarrow N^-} \frac{1}{N} \ln e = \frac{1}{N}$$

o mesmo resultado obtido para o caso de $t \rightarrow N^+$. Isso conclui a demonstração da regra da função logarítmica. Mais uma vez, note que, no processo de demonstração, nenhum valor numérico específico é empregado e, por conseguinte, o resultado é aplicável de modo geral.

Regra da função exponencial

A derivada da função $y = e^t$ é

$$\frac{d}{dt} e^t = e^t$$

Esse resultado é facilmente deduzido da regra da função logarítmica. Sabemos que a função inversa da função $y = e^t$ é $t = \ln y$, cuja derivada é $dt/dy = 1/y$. Assim, pela regra da função inversa, podemos escrever imediatamente

$$\frac{d}{dt} e^t = \frac{dy}{dt} = \frac{1}{dt/dy} = \frac{1}{1/y} = y = e^t$$

As regras generalizadas

As regras das funções logarítmica e exponencial podem ser generalizadas para os casos em que a variável t nas expressões e^t e $\ln t$ é substituída por alguma função de t , digamos, $f(t)$. As versões generalizadas das duas regras são

$$\begin{aligned} \frac{d}{dt} e^{f(t)} &= f'(t) e^{f(t)} & \left[\text{ou } \frac{d}{dt} e^u &= e^u \frac{du}{dt} \right] \\ \frac{d}{dt} \ln f(t) &= \frac{f'(t)}{f(t)} & \left[\text{ou } \frac{d}{dt} \ln v &= \frac{1}{v} \frac{dv}{dt} \right] \end{aligned} \quad (10.20)$$

As demonstrações para (10.20) são aplicações diretas da regra da cadeia. Dada uma função $y = e^{f(t)}$, podemos primeiro fazer $u = f(t)$, de modo que $y = e^u$. Então, pela regra da cadeia, a derivada surge como

$$\frac{d}{dt} e^{f(t)} = \frac{d}{dt} e^u = \frac{de^u}{du} \frac{du}{dt} = e^u \frac{du}{dt} = e^{f(t)} f'(t)$$

Da mesma forma, dada uma função $y = \ln f(t)$, podemos primeiro fazer $v = f(t)$, de modo a formar uma cadeia: $y = \ln v$, onde $v = f(t)$. Então, pela regra da cadeia, temos

$$\frac{d}{dt} \ln f(t) = \frac{d}{dt} \ln v = \frac{d \ln v}{dv} \frac{dv}{dt} = \frac{1}{v} \frac{dv}{dt} = \frac{1}{f(t)} f'(t)$$

Note que a única modificação real introduzida em (10.20), além das regras mais simples $de^t/dt = e^t$ e $d(\ln t)/dt = 1/t$ é o fator multiplicativo $f'(t)$.

EXEMPLO 1 Calcule a derivada da função $y = e^{rt}$. Aqui, o expoente é $rt = f(t)$, sendo $f'(t) = r$; assim,

$$\frac{dy}{dt} = \frac{d}{dt} e^{rt} = re^{rt}$$

EXEMPLO 2 Calcule dy/dt da função $y = e^{-t}$. Nesse caso, $f(t) = -t$, de modo que $f'(t) = -1$. Conseqüentemente,

$$\frac{dy}{dt} = \frac{d}{dt} e^{-t} = -e^{-t}$$

EXEMPLO 3 Calcule dy/dt da função $y = \ln(at)$. Uma vez que, nesse caso, $f(t) = at$, sendo $f'(t) = a$, a derivada é

$$\frac{d}{dt} \ln(at) = \frac{a}{at} = \frac{1}{t}$$

a qual, é interessante notar, é idêntica à derivada de $y = \ln t$.

Este exemplo ilustra o fato de que uma constante multiplicativa para t dentro de uma expressão logarítmica é descartada no processo de derivação. Mas note que, para uma constante k , temos

$$\frac{d}{dt} k \ln t = k \frac{d}{dt} \ln t = \frac{k}{t}$$

assim, uma constante multiplicativa fora da expressão logarítmica ainda é retida na derivação.

EXEMPLO 4 Calcule a derivada da função $y = \ln t^c$. Sendo $f(t) = t^c$ e $f'(t) = ct^{c-1}$, a fórmula em (10.20) dá

$$\frac{d}{dt} \ln t^c = \frac{ct^{c-1}}{t^c} = \frac{c}{t}$$

EXEMPLO 5 Calcule dy/dt de $y = t^3 \ln t^2$. Como essa função é um produto de dois termos, t^3 e $\ln t^2$, a regra do produto deve ser usada:

$$\begin{aligned} \frac{dy}{dt} &= t^3 \frac{d}{dt} \ln t^2 + (\ln t^2) \frac{d}{dt} t^3 \\ &= t^3 \left(\frac{2t}{t^2} \right) + (\ln t^2)(3t^2) \\ &= 2t^2 + 3t^2(2 \ln t) \quad [\text{Regra III de logaritmos}] \\ &= 2t^2(1 + 3 \ln t) \end{aligned}$$

O caso da base b

Para funções exponenciais e logarítmicas de base b , as derivadas são

$$\frac{d}{dt} b^t = b^t \ln b \quad \left[\text{Atenção: } \frac{d}{dt} b^t \neq t b^{t-1} \right] \quad (10.21)$$

$$\frac{d}{dt} \log_b t = \frac{1}{t \ln b}$$

Note que, no caso especial da base e (quando $b = e$), temos $\ln b = \ln e = 1$, de modo que essas duas derivadas se reduzem à regra básica da função exponencial $(d/dt) e^t = e^t$ e à regra básica da função logarítmica $(d/dt) \ln t = 1/t$, respectivamente.

A demonstração das fórmulas em (10.21) não é difícil. Para o caso de b^t , a demonstração é baseada na identidade $b \equiv e^{\ln b}$, que nos habilita a escrever

$$b^t = e^{(\ln b)t} = e^{t \ln b}$$

(Escrevemos $t \ln b$, em vez de $\ln bt$, para enfatizar que t não é uma parte da expressão logarítmica.) Portanto,

$$\begin{aligned} \frac{d}{dt} b^t &= \frac{d}{dt} e^{t \ln b} = (\ln b) (e^{t \ln b}) \quad [\text{por (10.20)}] \\ &= (\ln b) (b^t) = b^t \ln b \end{aligned}$$

Para provar a segunda parte de (10.21), por outro lado, recorreremos à propriedade básica dos logaritmos

$$\log_b t = (\log_b e) (\log_e t) = \frac{1}{\ln b} \ln t$$

que nos leva à derivada

$$\frac{d}{dt} \log_b t = \frac{d}{dt} \left(\frac{1}{\ln b} \ln t \right) = \frac{1}{\ln b} \frac{d}{dt} \ln t = \frac{1}{\ln b} \left(\frac{1}{t} \right)$$

As versões mais gerais dessas duas fórmulas são

$$\frac{d}{dt} b^{f(t)} = f'(t) b^{f(t)} \ln b \quad (10.21')$$

$$\frac{d}{dt} \log_b f(t) = \frac{f'(t)}{f(t)} \frac{1}{\ln b}$$

Mais uma vez, vemos que, se $b = e$, então $\ln b = 1$, e essas fórmulas se reduzem a (10.20).

EXEMPLO 6 Calcule a derivada da função $y = 12^{1-t}$. Aqui, $b = 12$; $f(t) = 1 - t$; e $f'(t) = -1$; assim,

$$\frac{dy}{dt} = -(12)^{1-t} \ln 12$$

Derivadas de ordens mais altas

Derivadas de ordens mais altas de funções exponenciais e logarítmicas, assim como derivadas de outros tipos de funções, são meros resultados de diferenciação repetida.

**EXEMPLO 7**

Calcule a derivada *segunda* de $y = b^t$ (com $b > 1$). A derivada primeira, por (10.21), é $y'(t) = b^t \ln b$ (onde $\ln b$ é, obviamente, uma constante); assim, diferenciando mais uma vez em relação a t , temos

$$y''(t) = \frac{d}{dt} y'(t) = \left(\frac{d}{dt} b^t \right) \ln b = (b^t \ln b) \ln b = b^t (\ln b)^2$$

Note que $y = b^t$ é sempre positivo e $\ln b$ (para $b > 1$) também é positivo [por (10.16)]; assim, $y'(t) = b^t \ln b$ deve ser positiva. E $y''(t)$, sendo um produto de b^t e um número elevado ao quadrado, também é positiva. Esses fatos confirmam o que afirmamos anteriormente – que a função exponencial $y = b^t$ aumenta monotonicamente a uma taxa crescente.

EXEMPLO 8

Calcule a derivada *segunda* de $y = \ln t$. A derivada primeira é $y' = 1/t = t^{-1}$; portanto, a derivada segunda é

$$y'' = -t^{-2} = \frac{-1}{t^2}$$

Considerando que o domínio dessa função consiste no intervalo aberto $(0, \infty)$, $y' = 1/t$ deve ser um número positivo. Por outro lado, y'' é sempre negativa. Juntas, essas conclusões servem para reiterar o que afirmamos antes: a função logarítmica $y = \ln t$ aumenta monotonicamente a uma taxa decrescente.

Uma aplicação

Uma das principais virtudes do logaritmo é sua capacidade de converter uma multiplicação em uma adição, e uma divisão em uma subtração. Essa propriedade pode ser explorada quando estamos diferenciando um *produto* ou um *quociente* complicado de qualquer tipo de função (não necessariamente exponencial ou logarítmica).

EXEMPLO 9

Calcule dy/dx onde

$$y = \frac{x^2}{(x+3)(2x+1)}$$

Em vez de aplicar as regras do quociente e do produto, podemos primeiramente tomar o log natural de ambos os lados da equação para reduzir a função à forma

$$\ln y = \ln x^2 - \ln(x+3) - \ln(2x+1)$$

De acordo com (10.20), a derivada do lado esquerdo em relação a x é

$$\frac{d}{dx} (\text{lado esquerdo}) = \frac{1}{y} \frac{dy}{dx}$$

enquanto o lado direito dá

$$\frac{d}{dx} (\text{lado direito}) = \frac{2x}{x^2} - \frac{1}{x+3} - \frac{2}{2x+1} = \frac{7x+6}{x(x+3)(2x+1)}$$

Quando igualamos os dois resultados e multiplicamos ambos os lados por y , obtemos a derivada desejada como se segue:

$$\begin{aligned} \frac{dy}{dx} &= \frac{7x+6}{x(x+3)(2x+1)} y \\ &= \frac{7x+6}{x(x+3)(2x+1)} \frac{x^2}{(x+3)(2x+1)} = \frac{x(7x+6)}{(x+3)^2(2x+1)^2} \end{aligned}$$

EXEMPLO 10

Calcule dy/dx onde $y = x^a e^{kx-c}$. Tomando o log natural de ambos os lados, temos

$$\ln y = a \ln x + \ln e^{kx-c} = a \ln x + kx - c$$

Diferenciando ambos os lados em relação a x , e usando (10.20), obtemos

$$\frac{1}{y} \frac{dy}{dx} = \frac{a}{x} + k$$

e

$$\frac{dy}{dx} = \left(\frac{a}{x} + k \right) y = \left(\frac{a}{x} + k \right) x^a e^{kx-c}$$

Contudo, note que, se a função dada contiver termos aditivos, pode *não* ser desejável converter a função para a forma logarítmica.

EXERCÍCIO 10.5

1. Calcule as derivadas de:

(a) $y = e^{2t+4}$

(d) $y = 5e^{2-t^2}$

(g) $y = x^2 e^{2x}$

(b) $y = e^{1-9t}$

(e) $y = e^{ax^2+bx+c}$

(h) $y = ax e^{bx+c}$

(c) $y = e^{t^2+1}$

(f) $y = xe^x$

2. (a) Verifique a derivada no Exemplo 3 utilizando a equação $\ln(at) = \ln a + \ln t$.

(b) Verifique o resultado no Exemplo 4 utilizando a equação $\ln t^c = c \ln t$.

3. Calcule as derivadas de:

(a) $y = \ln(7t^5)$

(d) $y = 5 \ln(t+1)^2$

(g) $y = \ln\left(\frac{2x}{1+x}\right)$

(b) $y = \ln(at^c)$

(e) $y = \ln x - \ln(1+x)$

(h) $y = 5x^4 \ln x^2$

(c) $y = \ln(t+19)$

(f) $y = \ln[x(1-x)^8]$

4. Calcule as derivadas de:

(a) $y = 5^t$

(c) $y = 13^{2t+3}$

(e) $y = \log_2(8x^2 + 3)$

(b) $y = \log_2(t+1)$

(d) $y = \log_7 7x^2$

(f) $y = x^2 \log_3 x$

5. Demonstre as duas fórmulas em (10.21').

6. Mostre que a função $V = Ae^{rt}$ (com $A, r > 0$) e a função $A = Ve^{-rt}$ (com $V, r > 0$) são ambas estritamente monotônicas, mas em direções opostas, e que a forma de ambas é estritamente convexa (ver Exercício 10.2-5).

7. Calcule as derivadas das seguintes funções, tomando, em primeiro lugar, o log natural de ambos os lados:

(a) $y = \frac{3x}{(x+2)(x+4)}$

(b) $y = (x^2 + 3)e^{x^2+1}$

10.6 Tempo ótimo

O que já aprendemos sobre funções exponenciais e logarítmicas agora pode ser aplicado a alguns problemas simples de tempo ótimo.

Um problema de armazenagem de vinho

Suponha que um certo comerciante de vinhos possui uma quantidade específica (digamos, uma caixa) de vinho que ele pode vender no tempo presente ($t=0$) por uma soma de $\$K$ ou armazenar por algum tempo e depois vender por um valor mais alto. Sabe-se que o valor crescente (V) do vinho é a seguinte função do tempo:

$$V = Ke^{\sqrt{t}} \quad [= K \exp(t^{1/2})] \quad (10.22)$$

de modo que, se $t = 0$ (vender agora), então $V = K$. O problema é determinar quando vendê-lo de modo a maximizar o lucro, admitindo que o custo de armazenagem seja nulo.[†]

Uma vez que o custo do vinho é um custo submerso ("sunk") – ele já foi pago pelo comerciante – e já que admitimos não existir custo de armazenagem, maximizar o lucro é o mesmo que maximizar a receita de vendas, ou o valor de V . Há, porém, um senão. Cada valor de V correspondente a um ponto específico de t representa uma soma em dólares a ser recebida em uma data diferente e, devido ao elemento juro envolvido, não pode ser comparado diretamente com o valor de V em uma outra data. O modo de sair dessa dificuldade é *descontar* cada valor V para seu *valor presente* equivalente (o valor no tempo $t = 0$) pois, então, todos os valores de V estarão no mesmo nível de comparação.

Vamos admitir que a taxa de juro composto contínuo esteja no nível de r . Então, segundo (10.11), o valor presente de V pode ser expresso como

$$A(t) = Ve^{-rt} = Ke^{\sqrt{t}} e^{-rt} = Ke^{\sqrt{t} - rt} \quad (10.22')$$

onde A , que denota o valor presente de V , é, em si, uma função de t . Por conseguinte, nosso problema trata de achar o valor de t que maximize A .

Condições de maximização

A condição de primeira ordem para maximizar A é ter $dA/dt = 0$. Para achar essa derivada, podemos ou diferenciar (10.22') diretamente em relação a t , ou fazer isso indiretamente primeiro tomando o logaritmo natural de ambos os lados de (10.22') e então diferenciando em relação a t . Vamos ilustrar o último procedimento.

Primeiro, obtemos de (10.22') a equação

$$\ln A(t) = \ln K + \ln e^{\sqrt{t} - rt} = \ln K + (t^{1/2} - rt)$$

Diferenciando ambos os lados em relação a t , obtemos então

$$\frac{1}{A} \frac{dA}{dt} = \frac{1}{2} t^{-1/2} - r$$

ou

$$\frac{dA}{dt} = A \left(\frac{1}{2} t^{-1/2} - r \right)$$

Visto que $A \neq 0$, a condição $dA/dt = 0$ pode ser satisfeita se e somente se

$$\frac{1}{2} t^{-1/2} = r \quad \text{ou} \quad \frac{1}{2\sqrt{t}} = r \quad \text{ou} \quad \frac{1}{2r} = \sqrt{t}$$

Isso implica que o período de tempo de armazenagem ótimo é

$$t^* = \left(\frac{1}{2r} \right)^2 = \frac{1}{4r^2}$$

Se $r = 0,10$, por exemplo, então $t^* = 25$, e o comerciante deverá armazenar a caixa de vinho por 25 anos. Note que, quanto mais alta a taxa de juro (taxa de desconto), menor será o período de armazenagem ótimo.

[†] Considerar o custo de armazenagem acarretará uma dificuldade que ainda não estamos aptos a enfrentar. Mais adiante, no Capítulo 14, voltaremos a esse problema.

A condição de primeira ordem, $1/(2\sqrt{t}) = r$, admite uma interpretação econômica fácil. A expressão do lado esquerdo representa tão somente a taxa de crescimento do valor do vinho V , porque, de (10.22)

$$\frac{dV}{dt} = \frac{d}{dt} K \exp(t^{1/2}) = K \frac{d}{dt} \exp(t^{1/2}) \quad [K \text{ constante}]$$

$$= K \left(\frac{1}{2} t^{-1/2} \right) \exp(t^{1/2}) \quad [\text{por (10.20)}]$$

$$= \left(\frac{1}{2} t^{-1/2} \right) V \quad [\text{por (10.22)}]$$

de modo que a taxa de crescimento de V é, de fato, a expressão do lado esquerdo da condição de primeira ordem:

$$r_V \equiv \frac{dV/dt}{V} = \frac{1}{2} t^{-1/2} = \frac{1}{2\sqrt{t}}$$

A expressão do lado direito, r , é, ao contrário, a taxa de juros ou a taxa de crescimento de juro composto dos fundos em caixa que receberiam se o vinho fosse vendido imediatamente – um aspecto de *custo de oportunidade* de armazenar o vinho. Assim, igualar as duas taxas instantâneas, como ilustrado na Figura 10.4, é uma tentativa de segurar o vinho até que a vantagem de armazená-lo desapareça totalmente, isto é, esperar até o exato momento em que a taxa de crescimento (decrecente) do valor do vinho torna-se igual à taxa de juro (constante) sobre as receitas de vendas.

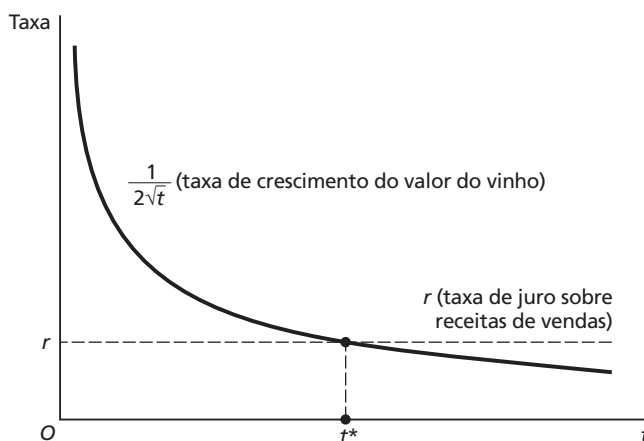
A etapa seguinte é verificar se o valor de t^* satisfaz a condição de segunda ordem para a maximização de A . A derivada segunda de A é

$$\frac{d^2 A}{dt^2} = \frac{d}{dt} A \left(\frac{1}{2} t^{-1/2} - r \right) = A \frac{d}{dt} \left(\frac{1}{2} t^{-1/2} - r \right) + \left(\frac{1}{2} t^{-1/2} - r \right) \frac{dA}{dt}$$

Mas, visto que o termo final é descartado quando o avaliamos no ponto de equilíbrio (ótimo), onde $dA/dt = 0$, ficamos com

$$\frac{d^2 A}{dt^2} = A \frac{d}{dt} \left(\frac{1}{2} t^{-1/2} - r \right) = A \left(-\frac{1}{4} t^{-3/2} \right) = \frac{-A}{4\sqrt{t^3}}$$

FIGURA 10.4



Em vista de $A > 0$, essa derivada segunda é negativa quando avaliada em $t^* > 0$, e, por isso, assegura que o valor de solução de t^* é, de fato, maximizador do lucro.

Um problema de corte de madeira

Um problema semelhante, que envolve a escolha da melhor época para entrar em ação, é o corte de madeira.

Suponha que o valor da madeira (árvores já plantadas em algum terreno dado) é a seguinte função crescente do tempo:

$$V = 2^{\sqrt{t}}$$

expressa em unidades de \$1.000. Admitindo uma taxa de desconto de r (em base contínua) e também admitindo custo zero de manutenção durante o período de crescimento das árvores, qual é o tempo ótimo para cortá-las para venda?

Como no problema do vinho, em primeiro lugar, devemos converter V a seu valor presente:

$$A(t) = Ve^{-rt} = 2^{\sqrt{t}} e^{-rt}$$

assim,
$$\ln A = \ln 2^{\sqrt{t}} + \ln e^{-rt} = \sqrt{t} \ln 2 - rt = t^{1/2} \ln 2 - rt$$

Para maximizar A , devemos estabelecer $dA/dt = 0$. A derivada primeira é obtida diferenciando $\ln A$ em relação a t e depois multiplicando por A :

$$\frac{1}{A} \frac{dA}{dt} = \frac{1}{2} t^{-1/2} \ln 2 - r$$

assim,
$$\frac{dA}{dt} = A \left(\frac{\ln 2}{2\sqrt{t}} - r \right)$$

Visto que $A \neq 0$, a condição $dA/dt = 0$ pode ser satisfeita se e somente se

$$\frac{\ln 2}{2\sqrt{t}} = r \quad \text{ou} \quad \sqrt{t} = \frac{\ln 2}{2r}$$

Por consequência, o número ótimo de anos de crescimento é

$$t^* = \left(\frac{\ln 2}{2r} \right)^2$$

Fica evidente por essa solução que, quanto mais alta a taxa de desconto, mais cedo a madeira deve ser cortada.

Para ter certeza de que t^* é uma solução maximizadora (em vez de minimizadora), a condição de segunda ordem deve ser verificada. Mas deixaremos que você faça isso como exercício.

Neste exemplo, abstraímos o custo de plantio, admitindo que as árvores já estão plantadas, caso em que é válido excluir tal custo (submerso) ao considerar a decisão de otimização. Se a decisão não se referir a quando cortar, mas a plantar ou não, então o custo do plantio (incorrido no *presente*) deve ser devidamente comparado com o valor *presente* da produção de madeira, em cujo cálculo t deve corresponder ao valor ótimo t^* . Por exemplo, se $r = 0,05$, então temos

$$t^* = \left(\frac{0,6931}{0,10} \right)^2 = (6,931)^2 = 48 \text{ anos}$$

e
$$A^* = 2^{6,931} e^{-0,05(48)} = (122,0222) e^{-2,40}$$

$$= 122,0222(0,0907) = \$11,0674 \text{ (em milhares)}$$

Portanto, somente um custo de plantio menor que A^* fará com que o empreendimento seja rentável – e, novamente, contanto que o custo de manutenção seja nulo.

EXERCÍCIO 10.6

1. Se o valor do vinho crescer segundo a função $V = K e^{2\sqrt{t}}$, em vez de (10.22), por quanto tempo o comerciante deve armazenar o vinho?
2. Verifique a condição de segunda ordem para o problema do corte de madeira.
3. Como uma generalização do problema de otimização ilustrado nesta seção, mostre que:
 - (a) Com qualquer valor da função $V = f(t)$ e uma dada taxa contínua de desconto r , a condição de primeira ordem para que o valor presente A de V alcance um máximo é que a taxa de crescimento de V seja igual a r .
 - (b) A condição suficiente de segunda ordem para um máximo realmente equivale a estipular que a taxa de crescimento de V será estritamente decrescente ao longo do tempo.
4. Analise a estática comparativa do problema da armazenagem do vinho.

10.7 Outras aplicações de derivadas exponenciais e logarítmicas

À parte sua utilização em problemas de otimização, as fórmulas de derivadas da Seção 10.5 têm outras aplicações úteis em economia.

Achando a taxa de crescimento

Quando uma variável y é uma função de tempo, $y = f(t)$, sua taxa instantânea de crescimento é definida como[†]

$$r_y \equiv \frac{dy/dt}{y} = \frac{f'(t)}{f(t)} = \frac{\text{função marginal}}{\text{função total}} \quad (10.23)$$

Mas, examinando (10.20), vemos que essa razão é precisamente a derivada de $\ln f(t) = \ln y$. Assim, para achar a taxa instantânea de crescimento de uma função de tempo $f(t)$, podemos – em vez de diferenciá-la em relação a t , e depois dividi-la por $f(t)$ – simplesmente tomar seu log natural e então diferenciar $\ln f(t)$ em relação ao tempo.^{††} Esse método alternativo pode ser uma abordagem mais simples se $f(t)$ for uma expressão de divisão ou de multiplicação pois, tomando o logaritmo dessas expressões, elas se reduzem a uma soma ou diferença de termos aditivos.

EXEMPLO 1

Calcule a taxa de crescimento de $V = Ae^{rt}$, onde t denota tempo. Já sabemos que a taxa de crescimento de V é r , mas vamos verificar, achando a derivada de $\ln V$:

$$\ln V = \ln A + rt \quad \text{e} \quad \ln e = \ln A + rt \quad [\text{A constante}]$$

Por conseguinte,

$$r_V = \frac{d}{dt} \ln V = 0 + \frac{d}{dt} rt = r$$

como queríamos demonstrar.

[†] Se a variável t não denotar tempo, a expressão $(dy/dt)/y$ é denominada *taxa proporcional de variação* de y em relação a t .

^{††} Se traçarmos o log natural de uma função $f(t)$ em relação a t em um diagrama bidimensional, a inclinação da curva nos dará a taxa de crescimento de $f(t)$. Isso nos fornece o princípio racional para as denominadas tábuas semilogarítmicas, que são usadas para comparar as taxas de crescimento de variáveis diferentes, ou as taxas de crescimento da mesma variável em países diferentes.

EXEMPLO 2 Calcule a taxa de crescimento de $y = 4^t$. Neste caso, temos

$$\ln y = \ln 4^t = t \ln 4$$

Portanto,

$$r_y = \frac{d}{dt} \ln y = \ln 4$$

E é assim mesmo que deve ser, porque $e^{\ln 4} = 4$ e, por consequência, $y = 4^t$ pode ser reescrita como $y = e^{(\ln 4)t}$, que nos habilitaria imediatamente a ler $(\ln 4)$ como a taxa de crescimento de y .

Taxa de crescimento de uma combinação de funções

Para levar essa discussão um pouco mais adiante, vamos examinar a taxa instantânea de crescimento de um *produto* de duas funções de tempo:

$$y = uv \quad \text{onde} \quad \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

Tomando o log natural de y , obtemos

$$\ln y = \ln u + \ln v$$

Assim, a taxa de crescimento desejada é

$$r_y = \frac{d}{dt} \ln y = \frac{d}{dt} \ln u + \frac{d}{dt} \ln v$$

Mas os dois termos do lado direito são as taxas de crescimento de u e v , respectivamente. Assim, temos a regra

$$r_{(uv)} = r_u + r_v \quad (10.24)$$

Expressa em palavras, a taxa instantânea de crescimento de um *produto* é a *soma* das taxas instantâneas de crescimento dos componentes.

Por um procedimento semelhante, pode-se demonstrar que a taxa de crescimento de um *quociente* é a *diferença* entre as taxas de crescimento dos componentes (veja Exercício 10.7-4):

$$r_{(u/v)} = r_u - r_v \quad (10.25)$$

EXEMPLO 3 Se o consumo C estiver crescendo a uma taxa α , e se a população H [H de “heads” (cabeças)] estiver crescendo à taxa β , qual é a taxa de crescimento do consumo *per capita*? Visto que o consumo *per capita* é igual a C/H , sua taxa de crescimento deve ser

$$r_{(C/H)} = r_C - r_H = \alpha - \beta$$

Agora considere a taxa instantânea de crescimento de uma *soma* de duas funções de tempo:

$$z = u + v \quad \text{onde} \quad \begin{cases} u = f(t) \\ v = g(t) \end{cases}$$

Dessa vez, o logaritmo natural será

$$\ln z = \ln(u + v) \quad [\neq \ln u + \ln v]$$

Assim,

$$\begin{aligned} r_z &= \frac{d}{dt} \ln z = \frac{d}{dt} \ln(u+v) \\ &= \frac{1}{u+v} \frac{d}{dt} (u+v) \quad [\text{por (10.20)}] \\ &= \frac{1}{u+v} [f'(t) + g'(t)] \end{aligned}$$

Mas de (10.23) temos $r_u = f'(t)/f(t)$, de modo que $f'(t) = f(t)r_u = ur_u$. De modo semelhante, temos $g'(t) = vr_v$. Como resultado, podemos escrever a regra

$$r_{(u+v)} = \frac{u}{u+v} r_u + \frac{v}{u+v} r_v \quad (10.26)$$

que mostra que a taxa de crescimento de uma soma é uma *média ponderada* das taxas de crescimento dos componentes.

Pelo mesmo critério, temos (veja Exercício 10.7-5)

$$r_{(u-v)} = \frac{u}{u-v} r_u - \frac{v}{u-v} r_v \quad (10.27)$$

EXEMPLO 4

As exportações de bens de um país, $G = G(t)$, têm uma taxa de crescimento a/t , e as exportações de seus serviços, $S = S(t)$, têm uma taxa de crescimento de b/t . Qual é a taxa de crescimento do total de suas exportações? Visto que o total de exportações é $X(t) = G(t) + S(t)$, uma soma, sua taxa de crescimento deve ser

$$\begin{aligned} r_X &= \frac{G}{X} r_G + \frac{S}{X} r_S \\ &= \frac{G}{X} \left(\frac{a}{t} \right) + \frac{S}{X} \left(\frac{b}{t} \right) = \frac{Ga + Sb}{Xt} \end{aligned}$$

Achando a elasticidade pontual

Vimos que, dada $y = f(t)$, a derivada de $\ln y$ mede a taxa instantânea de crescimento de y . Agora, vamos ver o que acontece quando, dada uma função $y = f(x)$, diferenciamos $\ln y$ em relação a $\ln x$, em vez de a x .

Para começar, vamos definir $u \equiv \ln y$ e $v \equiv \ln x$. Então, podemos observar uma cadeia de relações que vinculam u a y , e, por conseguinte, x a v , como se segue:

$$u \equiv \ln y \quad y = f(x) \quad x \equiv e^{\ln x} \equiv e^v$$

De acordo com isso, a derivada de $\ln y$ em relação a $\ln x$ é

$$\begin{aligned} \frac{d(\ln y)}{d(\ln x)} &= \frac{du}{dv} = \frac{du}{dy} \frac{dy}{dx} \frac{dx}{dv} \\ &= \left(\frac{d}{dy} \ln y \right) \left(\frac{dy}{dx} \right) \left(\frac{d}{dv} e^v \right) = \frac{1}{y} \frac{dy}{dx} e^v = \frac{1}{y} \frac{dy}{dx} x = \frac{dy}{dx} \frac{x}{y} \end{aligned}$$

Mas essa expressão é exatamente a expressão da elasticidade pontual da função. Por conseguinte, estabelecemos o princípio geral de que, para uma função $y = f(x)$, a elasticidade pontual de y em relação a x é

$$\varepsilon_{yx} = \frac{d(\ln y)}{d(\ln x)} \quad (10.28)$$

Deve-se notar que o índice yx nesse símbolo é um indicador de que y e x são as duas variáveis envolvidas e não implica a multiplicação de y e x . Isso é diferente do caso de $r_{(uv)}$, onde o índice *realmente* denota um produto. Novamente, agora temos um modo alternativo de achar a elasticidade pontual de uma função com a utilização de logaritmos, o que muitas vezes pode se revelar uma abordagem mais fácil, se a função dada vier na forma de uma expressão de multiplicação ou de divisão.

EXEMPLO 5

Encontre a elasticidade pontual de demanda, dado que $Q = k/P$, onde k é uma constante positiva. Essa é a equação de uma hipérbole retangular (veja Figura 2.8d); e, como sabemos muito bem, a elasticidade pontual de uma função demanda com essa forma é unitária em todos os pontos. Para demonstrar isso, aplicaremos (10.28). Posto que o logaritmo natural da função demanda é

$$\ln Q = \ln k - \ln P$$

a elasticidade de demanda (Q em relação a P) é, de fato,

$$\varepsilon_d = \frac{d(\ln Q)}{d(\ln P)} = -1 \quad \text{ou} \quad |\varepsilon_d| = 1$$

O resultado em (10.28) foi obtido com a utilização da regra da cadeia para derivadas. É de interesse que uma regra da cadeia semelhante seja válida para elasticidades; isto é, dada uma função $y = g(w)$, onde $w = h(x)$, temos

$$\varepsilon_{yx} = \varepsilon_{yw} \varepsilon_{wx} \quad (10.29)$$

A demonstração é a seguinte:

$$\varepsilon_{yw} \varepsilon_{wx} = \left(\frac{dy}{dw} \frac{w}{y} \right) \left(\frac{dw}{dx} \frac{x}{w} \right) = \frac{dy}{dw} \frac{dw}{dx} \frac{w}{y} \frac{x}{w} = \frac{dy}{dx} \frac{x}{y} = \varepsilon_{yx}$$

EXERCÍCIO 10.7

- Calcule a taxa instantânea de crescimento:
 - $y = 5t^2$
 - $y = at^c$
 - $y = ab^t$
 - $y = 2^t(t^2)$
 - $y = t/3^t$
- Se a população crescer segundo a função $H = H_0(2)^{bt}$ e o consumo segundo a função $C = C_0 e^{at}$, calcule as taxas de crescimento da população, do consumo e do consumo *per capita* usando o logaritmo natural.
- Se y for relacionado a x por $y = x^k$, como as taxas de crescimento r_y e r_x estarão relacionadas?
- Prove que, se $y = u/v$, onde $u = f(t)$ e $v = g(t)$, então a taxa de crescimento de y será $r_y = r_u - r_v$ como mostrado em (10.25).
- A receita real y é definida como a receita nominal Y deflacionada pelo nível de preço P . Como r_y (para renda real) está relacionada a r_Y (para renda nominal)?
- Prove a regra de taxa de crescimento (10.27).
- Dada a função demanda $Q_d = k/P^n$, onde k e n são constantes positivas, calcule a elasticidade pontual de demanda ε_d utilizando (10.28) (veja Exercício 8.1-4).
- Dada $y = wz$, onde $w = g(x)$ e $z = h(x)$, estabeleça que $\varepsilon_{yx} = \varepsilon_{wx} + \varepsilon_{zx}$.
 - Dada $y = u/v$, onde $u = G(x)$ e $v = H(x)$, estabeleça que $\varepsilon_{yx} = \varepsilon_{ux} - \varepsilon_{vx}$.
- Dada $y = f(x)$, mostre que a derivada $d(\log_b y)/d(\log_b x)$ – logaritmo na base b , e não na base e – também mede a elasticidade pontual ε_{yx} .

10. Mostre que, se a demanda por moeda M_d for uma função da renda nacional $Y = Y(t)$ e da taxa de juros $i = i(t)$, a taxa de crescimento de M_d pode ser expressa como uma soma ponderada de r_Y e r_i ,

$$r_{M_d} = \varepsilon_{M_d Y} r_Y + \varepsilon_{M_d i} r_i$$

onde os pesos são as elasticidades de M_d em relação a Y e i , respectivamente.

11. Dada a função produção $Q = F(K, L)$, encontre uma expressão geral para a taxa de crescimento de Q em termos das taxas de crescimento de K e L .

CAPÍTULO 11

O caso de mais de uma variável de escolha

O problema de otimização foi discutido no Capítulo 9, dentro da estrutura de uma função objetivo com uma única variável de escolha. No Capítulo 10, a discussão foi estendida para funções objetivo exponenciais, mas ainda tratávamos de uma única variável. Agora, temos de desenvolver um modo de achar os valores extremos de uma função objetivo que envolva duas ou mais variáveis de escolha. Somente então poderemos atacar o tipo de problema que enfrenta, por exemplo, uma empresa com vários produtos, onde a decisão de maximização do lucro consiste na escolha de níveis ótimos de produção para diversas mercadorias e na combinação ótima de diversos insumos diferentes.

Em primeiro lugar, discutiremos o caso de uma função objetivo com duas variáveis de escolha, $z = f(x, y)$, para aproveitar a vantagem da facilidade de representá-la graficamente. Mais adiante, os resultados analíticos podem ser generalizados para o caso de n variáveis, que não é possível representar graficamente. Independentemente do número de variáveis, entretanto, em termos gerais admitiremos que, quando escrita em uma forma geral, nossa função objetivo possui derivadas parciais contínuas até qualquer ordem desejada. Isso garantirá a suavidade e a diferenciabilidade da função objetivo, bem como de suas derivadas parciais.

Para funções com muitas variáveis, novamente os valores extremos são de dois tipos: (1) absolutos ou globais e (2) relativos ou locais. Como antes, focalizaremos nossa intensa atenção nos extremos relativos e, por essa razão, muitas vezes descartaremos o adjetivo “relativo”, entendendo que, se não houver especificação em contrário, os extremos a que nos referirmos são *relativos*. Contudo, na Seção 11.5, daremos a devida consideração aos extremos *absolutos*.

11.1 A versão diferencial de condições de otimização

A discussão no Capítulo 9, referente a condições de otimização para problemas com uma única variável de escolha, foi inteiramente expressa em termos de *derivadas*, em vez de diferenciais. Como preparação para a discussão de problemas com duas ou mais variáveis de escolha, seria útil também saber que essas condições podem ser expressas de modo equivalente em termos de *diferenciais*.

Condição de primeira ordem

Dada uma função $z = f(x)$, podemos, como explicado na Seção 8.1, escrever a diferencial

$$dz = f'(x) dx \quad (11.1)$$

e usar dz como uma aproximação da variação propriamente dita, Δz , induzida pela mudança em x de x_0 para $x_0 + \Delta x$; quanto menor o Δx , melhor é a aproximação. De (11.1) fica claro que, se $f'(x) > 0$, dz e dx devem assumir o mesmo sinal algébrico, o que é ilustrado pelo ponto A na Figura 11.1 (veja Figura 8.1b). No caso oposto em que $f'(x) < 0$, exemplificado pelo ponto A' , dz e dx assumem sinais algébricos opostos. Uma vez que pontos como A e A' – onde $f'(x) \neq 0$ e, por conseguinte, $dz \neq 0$ – não podem se qualificar como pontos estacionários, é óbvio que uma condição necessária para z atingir um extremo (um valor estacionário) é $dz = 0$. A condição deve ser enunciada, com maior exatidão, como “ $dz = 0$ para um dx diferente de zero arbitrário”, já que um dx igual a zero (nenhuma variação em x) não tem nenhuma relevância em nosso contexto presente. Na Figura 11.1, ocorre um mínimo de z no ponto B e um máximo de z no ponto B' . Em ambos os casos, como a reta tangente traçada nesses pontos é horizontal, isto é, $f'(x) = 0$, dz (que é o lado vertical do triângulo cuja hipotenusa é essa reta tangente) realmente se reduz a zero. Assim, a condição de *derivada* de primeira ordem “ $f'(x) = 0$ ” pode ser traduzida para a condição de *diferencial* de primeira ordem “ $dz = 0$ para um dx arbitrário diferente de zero.” Porém, não se esqueça que, conquanto essa condição diferencial seja necessária para um extremo, ela não é, de modo algum, suficiente, porque um ponto de inflexão, tal como C na Figura 11.1, também pode satisfazer a condição $dz = 0$ para um dx arbitrário diferente de zero.

Condição de segunda ordem

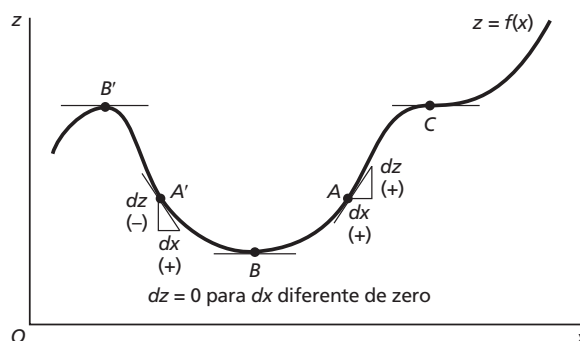
As condições suficientes de segunda ordem para extremos de z são, em termos de derivadas, $f''(x) < 0$ (para um máximo) e $f''(x) > 0$ (para um mínimo) no ponto estacionário. Para traduzir essas condições em diferenciais equivalentes, precisamos da noção de *diferencial de segunda ordem*, definida como a diferencial de uma diferencial, isto é, $d(dz)$, comumente denotada por d^2z (lê-se: “ d dois z ”).

Dado que $dz = f'(x) dx$, podemos obter d^2z por mera diferenciação ulterior de dz . Contudo, ao fazer isso, devemos ter em mente que dx , que nesse contexto representa uma dada variação diferente de zero em x , deve ser tratada como uma constante durante a diferenciação. Por consequência, dz pode variar somente com $f'(x)$, mas, posto que $f'(x)$ é, por sua vez, uma função de x , em última instância, dz varia somente com x . Em vista disso, temos

$$\begin{aligned} d^2z &\equiv d(dz) = d[f'(x) dx] && [\text{por (11.1)}] \\ &= [d f'(x)] dx && [dx \text{ é constante}] \\ &= [f''(x) dx] dx = f''(x) dx^2 \end{aligned} \quad (11.2)$$

Note que o expoente 2 aparece em (11.2) de dois modos fundamentalmente diferentes. No símbolo d^2z , o expoente 2 (lê-se: “dois”) indica a diferencial de *segunda ordem* de z ; mas no símbolo $dx^2 \equiv (dx)^2$, o expoente 2 (lê-se “ao quadrado”) denota a *elevação ao quadrado* da diferencial de primeira ordem dx . O resultado em (11.2) provê um vínculo direto entre d^2z e $f''(x)$. Como estamos considerando somente valores diferentes de zero de dx , o termo dx^2 é sempre positivo; assim, d^2z e $f''(x)$ devem assumir o mesmo sinal algébrico. Exatamente como uma $f''(x)$ positiva (negati-

FIGURA 11.1



va) em um ponto estacionário delineia um vale (pico), o mesmo deve acontecer com uma d^2z positiva (negativa) em tal ponto.

Segue-se que a condição de *derivada* “ $f''(x) < 0$ é suficiente para um máximo de z ” pode ser enunciada, de modo equivalente, como a condição de *diferencial* “ $d^2z < 0$ para um dx arbitrário diferente de zero é suficiente para um máximo de z ”. A tradução da condição para um mínimo de z é análoga; basta inverter o sentido da desigualdade na sentença precedente. Avançando um pouco mais, também podemos concluir, com base em (11.2), que as condições *necessárias* de segunda ordem são:

$$\begin{aligned} \text{Para o máximo de } z: & \quad f''(x) \leq 0 \\ \text{Para o mínimo de } z: & \quad f''(x) \geq 0 \end{aligned}$$

e podem ser traduzidas, respectivamente, em

$$\left. \begin{aligned} \text{Para o máximo de } z: & \quad d^2z \leq 0 \\ \text{Para o mínimo de } z: & \quad d^2z \geq 0 \end{aligned} \right\} \text{ para valores arbitrários diferentes de zero de } dx$$

Condições de diferencial *versus* condições de derivada

Agora que já demonstramos a possibilidade de expressar a versão derivada das condições de primeira e segunda ordem em termos de dz e d^2z , você pode muito bem perguntar por que nos demos ao trabalho de desenvolver um novo conjunto de condições de diferenciais quando já havia condições de derivadas disponíveis. A resposta é que as condições de diferenciais – mas não as condições de derivadas – são enunciadas de um modo tal que podem ser diretamente generalizadas do caso com uma variável para casos com duas ou mais variáveis de escolha. Para sermos mais específicos, a condição de primeira ordem (valor zero para dz) e a condição de segunda ordem (negatividade ou positividade para d^2z) são aplicáveis com igual validade a todos os casos, contanto que a frase “para valores de dx arbitrários diferentes de zero” seja devidamente modificada de modo a refletir a mudança no número de variáveis de escolha.

Entretanto, isso não significa que as condições de derivada já não terão mais nenhum papel a desempenhar. Ao contrário, visto que essas condições são mais convenientes de aplicar do ponto de vista operacional, ainda tentaremos – após executar o processo de generalização para casos com mais variáveis de escolha por meio das condições de diferencial – desenvolver e utilizar condições de derivada apropriadas para esses casos.

11.2 Valores extremos de uma função de duas variáveis

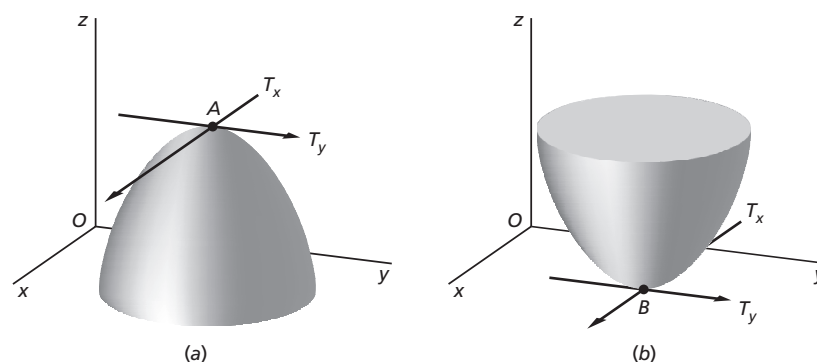
No caso de uma função de uma única variável, um valor extremo é representado graficamente como o pico de uma colina ou o fundo de um vale em um gráfico bidimensional. Com *duas* variáveis de escolha, o gráfico da função – $z = f(x, y)$ – torna-se uma superfície em um espaço tridimensional e, embora esses valores extremos ainda possam ser associados com picos e fundos, essas “colinas” e “vales” agora assumem um caráter tridimensional. Nesse novo contexto, as colinas e vales terão formatos de domos e tigelas, respectivamente. Os dois diagramas na Figura 11.2 servem como ilustração. O ponto A no digrama a , o pico de um domo, constitui um máximo; a valor de z nesse ponto é maior que em qualquer outro ponto em sua vizinhança imediata. De modo semelhante, o ponto B no diagrama b , o fundo de uma tigela, representa um mínimo; em todos os lugares de sua vizinhança imediata o valor da função é maior que no ponto B .

Condição de primeira ordem

Para a função

$$z = f(x, y)$$

FIGURA 11.2



a condição necessária de primeira ordem para um extremo (seja máximo ou mínimo) novamente envolve $dz = 0$. Mas, visto que aqui há duas variáveis independentes, dz agora é uma diferencial *total*; assim, a condição de primeira ordem deve ser modificada para a forma

$$dz = 0 \text{ para valores arbitrários de } dx \text{ e } dy, \text{ não sendo ambos iguais a zero} \quad (11.3)$$

O princípio racional por trás de (11.3) é semelhante à explicação da condição $dz = 0$ para o caso de uma só variável: um ponto extremo deve ser um ponto estacionário e, em um ponto estacionário, dz , na qualidade de uma aproximação da variação propriamente dita Δz , deve ser zero para dx e dy arbitrários, não sendo ambos iguais a zero.

No caso presente de duas variáveis, a diferencial total é

$$dz = f_x dx + f_y dy \quad (11.4)$$

Para satisfazer a condição (11.3), é necessário e suficiente que as duas derivadas parciais f_x e f_y sejam simultaneamente iguais a zero. Assim, a versão equivalente da condição de primeira ordem (11.3) é

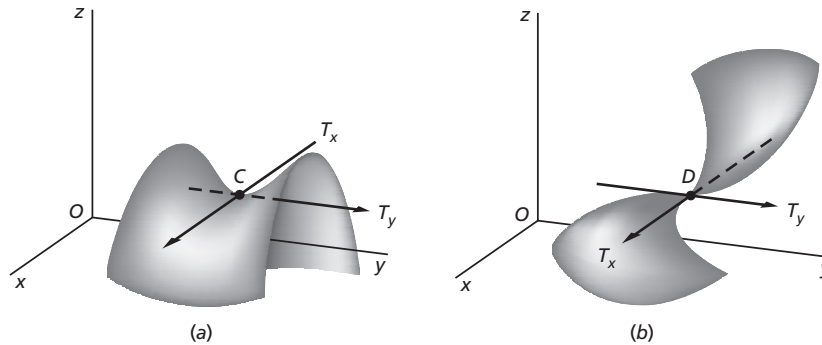
$$f_x = f_y = 0 \quad \left[\text{ou} \quad \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 0 \right] \quad (11.5)$$

Há uma interpretação gráfica simples dessa condição. Com referência ao ponto A na Figura 11.2a, ter $f_x = 0$ naquele ponto significa que a reta tangente T_x , desenhada no ponto A e paralela ao plano xz (mantendo y constante), deve ter uma inclinação igual a zero. Pelo mesmo critério, ter $f_y = 0$ no ponto A significa que a reta tangente T_y , desenhada no ponto A e paralela ao plano yz (mantendo x constante), também deve ter uma inclinação zero. Você pode verificar, de imediato, que esses requisitos da reta tangente na verdade também se aplicam ao ponto de mínimo B na Figura 11.2b. Isso porque a condição (11.5), assim como a condição (11.3), é uma condição necessária para *ambos*: um máximo e um mínimo.

Como na discussão anterior, a condição de primeira ordem é *necessária*, mas *não suficiente*. O fato de ela não ser suficiente para estabelecer um extremo pode ser visto nos dois diagramas na Figura 11.3. No ponto C do diagrama a, as inclinações de ambas as retas, T_x e T_y , são iguais a zero, mas esse ponto não se qualifica como um extremo: conquanto seja um *mínimo* quando considerado em relação ao plano yz , ele se transforma em um *máximo* quando considerado em relação ao plano xz ! Um ponto com tal “personalidade dupla” é denominado, por questões gráficas, um *ponto de sela*. De maneira semelhante, o ponto D na Figura 11.3b, embora caracterizado por T_x e T_y planas, tampouco é um extremo; sua localização na superfície retorcida o torna um *ponto de inflexão*, quer considerado em relação ao plano xz , quer em relação ao plano yz . Esses contra-exemplos decididamente eliminam a condição de primeira ordem como uma condição suficiente para um extremo.

Para desenvolver uma condição suficiente, temos de examinar a diferencial total de segunda ordem, que é relacionada com derivadas parciais de segunda ordem.

FIGURA 11.3



Derivadas parciais de segunda ordem

A função $z = f(x, y)$ pode dar origem a duas derivadas parciais de primeira ordem,

$$f_x \equiv \frac{\partial z}{\partial x} \quad \text{e} \quad f_y \equiv \frac{\partial z}{\partial y}$$

Visto que f_x é, em si, uma função de x (bem como de y), podemos medir a taxa de variação de f_x em relação a x , enquanto y permanece fixo, por uma derivada parcial de segunda ordem (ou derivada parcial segunda) específica, denotada ou por f_{xx} ou por $\partial^2 z / \partial x^2$:

$$f_{xx} \equiv \frac{\partial}{\partial x} (f_x) \quad \text{ou} \quad \frac{\partial^2 z}{\partial x^2} \equiv \frac{\partial}{\partial x} \left(\frac{\partial z}{\partial x} \right)$$

A notação f_{xx} tem um índice duplo que significa que a função primitiva f foi diferenciada parcialmente em relação a x duas vezes, ao passo que a notação $\partial^2 z / \partial x^2$ é parecida com a de $d^2 z / dx^2$ exceto pela utilização do símbolo de parcial. De maneira perfeitamente análoga, podemos usar a derivada parcial segunda

$$f_{yy} \equiv \frac{\partial}{\partial y} (f_y) \quad \text{ou} \quad \frac{\partial^2 z}{\partial y^2} \equiv \frac{\partial}{\partial y} \left(\frac{\partial z}{\partial y} \right)$$

para denotar a taxa de variação de f_y em relação a y , enquanto x for mantido constante.

Lembre-se, todavia, que f_x também é uma função de y e que f_y também é uma função de x . Portanto, podem ser escritas mais duas derivadas parciais segundas:

$$f_{xy} \equiv \frac{\partial^2 z}{\partial x \partial y} \equiv \frac{\partial}{\partial x} \left(\frac{\partial z}{\partial y} \right) \quad \text{ou} \quad f_{yx} \equiv \frac{\partial^2 z}{\partial y \partial x} \equiv \frac{\partial}{\partial y} \left(\frac{\partial z}{\partial x} \right)$$

Essas derivadas são denominadas *derivadas parciais cruzadas* (ou *mistas*) porque cada uma mede a taxa de variação de uma derivada parcial de primeira ordem em relação à “outra” variável.

Vale repetir que as derivadas parciais de segunda ordem de $z = f(x, y)$, assim como z e as derivadas primeiras f_x e f_y , também são funções das variáveis x e y . Quando for preciso destacar esse fato, podemos escrever f_{xx} como $f_{xx}(x, y)$, e f_{xy} como $f_{xy}(x, y)$ etc. E, pelo mesmo critério, podemos usar a notação $f_{yx}(1, 2)$ para denotar o valor de f_{yx} calculado em $x = 1$ e $y = 2$ etc.

Ainda que f_{xy} e f_{yx} tenham sido definidas em separado, elas terão – segundo uma proposição conhecida como *Teorema de Young* – valores idênticos, contanto que as duas derivadas parciais cruzadas sejam ambas contínuas. Nesse caso, a ordem de seqüência em que a diferenciação é realizada torna-se irrelevante, porque $f_{xy} = f_{yx}$. Para os tipos comuns de funções *específicas* com as quais trabalhamos, essa condição de continuidade usualmente é satisfeita; para funções *gerais*, como mencionado anteriormente, sempre admitimos que a condição de continuidade permanece válida. Por conseguinte, em geral podemos esperar encontrar derivadas parciais cruzadas idênticas. De fato, o teorema também se aplica a funções de três ou mais variáveis. Dada $z = g(u, v,$

$w)$, por exemplo, as derivadas parciais mistas serão caracterizadas por $g_{uv} = g_{vu}$, $g_{uvw} = g_{wuv}$ etc., contanto que essas derivadas parciais sejam todas contínuas.

EXEMPLO 1 Calcule as derivadas parciais de segunda ordem de

$$z = x^3 + 5xy - y^2$$

As derivadas parciais primeiras dessa função são

$$f_x = 3x^2 + 5y \quad \text{e} \quad f_y = 5x - 2y$$

Portanto, prosseguindo na diferenciação, obtemos

$$f_{xx} = 6x \quad f_{yx} = 5 \quad f_{xy} = 5 \quad f_{yy} = -2$$

Como esperado, f_{yx} e f_{xy} são idênticas.

EXEMPLO 2 Calcule todas as derivadas parciais segundas de $z = x^2 e^{-y}$. Nesse caso, as derivadas parciais primeiras são

$$f_x = 2xe^{-y} \quad \text{e} \quad f_y = -x^2 e^{-y}$$

Assim, temos

$$f_{xx} = 2e^{-y} \quad f_{yx} = -2xe^{-y} \quad f_{xy} = -2xe^{-y} \quad f_{yy} = x^2 e^{-y}$$

Mais uma vez, vemos que $f_{yx} = f_{xy}$.

Note que todas as derivadas parciais segundas são funções das variáveis originais x e y . Esse fato fica bastante claro no Exemplo 2, mas é verdadeiro até mesmo para o Exemplo 1, embora algumas derivadas parciais segundas por acaso sejam funções *constantes* naquele caso.

Diferencial total de segunda ordem

Dada a diferencial total dz em (11.4), e dominando o conceito de derivadas parciais de segunda ordem, podemos obter uma expressão para a diferencial total de segunda ordem d^2z prosseguindo na diferenciação de dz . Ao fazer isso, devemos nos lembrar que, na equação $dz = f_x dx + f_y dy$, os símbolos dx e dy representam variações arbitrárias ou definidas em x e y ; portanto, elas devem ser tratadas como constantes durante a diferenciação. Por consequência, dz depende somente de f_x e f_y e, visto que f_x e f_y são, em si, funções de x e y , dz , assim como a própria z , é uma função de x e y .

Para obter d^2z , basta aplicar a definição de uma diferencial – como mostrado em (11.4) – à própria dz . Assim,

$$\begin{aligned} d^2z &\equiv d(dz) = \frac{\partial(dz)}{\partial x} dx + \frac{\partial(dz)}{\partial y} dy \quad [\text{veja (11.4)}] \\ &= \frac{\partial}{\partial x} (f_x dx + f_y dy) dx + \frac{\partial}{\partial y} (f_x dx + f_y dy) dy \\ &= (f_{xx} dx + f_{xy} dy) dx + (f_{yx} dx + f_{yy} dy) dy \\ &= f_{xx} dx^2 + f_{xy} dy dx + f_{yx} dx dy + f_{yy} dy^2 \\ &= f_{xx} dx^2 + 2 f_{xy} dx dy + f_{yy} dy^2 \quad [f_{xy} = f_{yx}] \end{aligned} \quad (11.6)$$

Mais uma vez, note que o expoente 2 aparece em (11.6) de dois modos diferentes. No símbolo d^2z , o expoente 2 indica a diferencial total de *segunda ordem* de z ; mas, no símbolo $dx^2 \equiv (dx)^2$, o expoente denota a *elevação ao quadrado* da diferencial de primeira ordem dx .

O resultado em (11.6) mostra a grandeza de d^2z em termos de valores definidos de dx e dy , medida em relação a algum ponto (x_0, y_0) no domínio. Para calcular d^2z , contudo, também precisamos conhecer as derivadas parciais de segunda ordem f_{xx} , f_{xy} e f_{yy} , todas calculadas em (x_0, y_0) – assim como precisamos conhecer as derivadas parciais de primeira ordem para calcular dz a partir de (11.4).

EXEMPLO 3

Dada $z = x^3 + 5xy - y^2$, calcule dz e d^2z . Essa função é a mesma do Exemplo 1. Assim, substituindo as várias derivadas já obtidas em (11.4) e (11.6), obtemos[†]

$$dz = (3x^2 + 5y) dx + (5x - 2y) dy$$

e

$$d^2z = 6x dx^2 + 10 dx dy - 2 dy^2$$

Também podemos calcular dz e d^2z em pontos específicos no domínio. No ponto $x = 1$ e $y = 2$, por exemplo, temos

$$dz = 13 dx + dy \quad \text{e} \quad d^2z = 6 dx^2 + 10 dx dy - 2 dy^2$$

Condição de segunda ordem

No caso de uma só variável, $d^2z < 0$ em um ponto estacionário identifica o ponto como o pico de uma colina em um espaço bidimensional. De modo semelhante, no caso de duas variáveis, $d^2z < 0$ em um ponto estacionário identifica o ponto como o pico de um domo em um espaço tridimensional. Assim, uma vez satisfeita a condição necessária de primeira ordem, a condição suficiente de segunda ordem para um máximo de $z = f(x, y)$ é

$$d^2z < 0 \text{ para valores arbitrários } dx \text{ e } dy, \text{ não sendo ambos iguais a zero} \quad (11.7)$$

Uma d^2z de valor positivo em um ponto estacionário, por outro lado, é associada com o fundo de uma tigela. A condição suficiente de segunda ordem para um mínimo de $z = f(x, y)$ é

$$d^2z > 0 \text{ para valores arbitrários de } dx \text{ e } dy, \text{ não sendo ambos iguais a zero} \quad (11.8)$$

A razão por que (11.7) e (11.8) são condições apenas suficientes, mas não necessárias, é que, novamente, é possível que d^2z assuma um valor igual a zero em um máximo ou em um mínimo. Por essa razão, condições *necessárias* de segunda ordem devem ser enunciadas com desigualdades fracas, como se segue:

$$\left. \begin{array}{l} \text{Para o máximo de } z: \quad d^2z \leq 0 \\ \text{Para o mínimo de } z: \quad d^2z \geq 0 \end{array} \right\} \text{ para valores arbitrários de } dx \text{ e } dy, \text{ não sendo ambos iguais a zero.} \quad (11.9)$$

[†] Um modo alternativo de chegar a esses resultados é por diferenciação direta da função:

$$\begin{aligned} dz &= d(x^3) + d(5xy) - d(y^2) \\ &= 3x^2 dx + 5y dx + 5x dy - 2y dy \end{aligned}$$

Prosseguindo na diferenciação de dz (sem esquecer que dx e dy são constantes), teremos

$$\begin{aligned} d^2z &= d(3x^2) dx + d(5y) dx + d(5x) dy - d(2y) dy \\ &= (6x dx) dx + (5 dy) dx + (5 dx) dy - (2 dy) dy \\ &= 6x dx^2 + 10 dx dy - 2 dy^2 \end{aligned}$$

Entretanto, a seguir daremos mais atenção às condições suficientes de segunda ordem.

Por questão de conveniência operacional, condições de diferenciais de segunda ordem podem ser traduzidas em condições equivalentes de derivadas de segunda ordem. No caso de duas variáveis, (11.6) mostra que isso acarretaria restrições nos sinais das derivadas parciais de segunda ordem f_{xx} , f_{xy} e f_{yy} . A tradução propriamente dita exigiria o conhecimento de formas quadráticas, que serão discutidas na Seção 11.3. Porém, antes, podemos apresentar o resultado principal aqui: para quaisquer valores de dx e dy , não sendo ambos iguais a zero,

$$d^2z \begin{cases} < 0 & \text{se e somente se } f_{xx} < 0; \quad f_{yy} < 0; \quad \text{e } f_{xx}f_{yy} > f_{xy}^2 \\ > 0 & \text{se e somente se } f_{xx} > 0; \quad f_{yy} > 0; \quad \text{e } f_{xx}f_{yy} > f_{xy}^2 \end{cases}$$

Note que o sinal de d^2z depende não somente de f_{xx} e f_{yy} , que têm a ver com a configuração da superfície ao redor do ponto A (Figura 11.4) nas duas direções básicas mostradas por T_x (leste-oeste) e T_y (norte-sul), mas também da derivada parcial cruzada f_{xy} . O papel desempenhado por essa última derivada parcial é assegurar que a superfície em questão resultará em seções transversais (bidimensionais) com o mesmo tipo de configuração (colina ou vale, conforme o caso) não somente nas duas direções básicas (leste-oeste e norte-sul), mas também em todas as outras direções possíveis (tal como nordeste-sudoeste).

Esse resultado, juntamente com a condição de primeira ordem (11.5), nos habilita a montar a Tabela 11.1. É bom que se entenda que todas as derivadas parciais segundas ali presentes devem ser calculadas no ponto estacionário onde $f_x = f_y = 0$. É preciso também salientar que a condição suficiente de segunda ordem *não é necessária* para um extremo. Em particular, se um valor estacionário for caracterizado por $f_{xx}f_{yy} = f_{xy}^2$, violando aquela condição, esse valor estacionário pode, não obstante, resultar em um extremo. Por outro lado, no caso de um outro tipo de violação, com um ponto estacionário caracterizado por $f_{xx}f_{yy} < f_{xy}^2$, podemos identificar aquele ponto como um ponto de sela, porque o sinal de d^2z , nesse caso, será indefinido (positivo para alguns valores de dx e dy , mas negativo para outros).

EXEMPLO 4 Encontre o valor (ou valores) extremo de $z = 8x^3 + 2xy - 3x^2 + y^2 + 1$. Em primeiro lugar, vamos calcular todas as derivadas parciais primeiras e segundas:

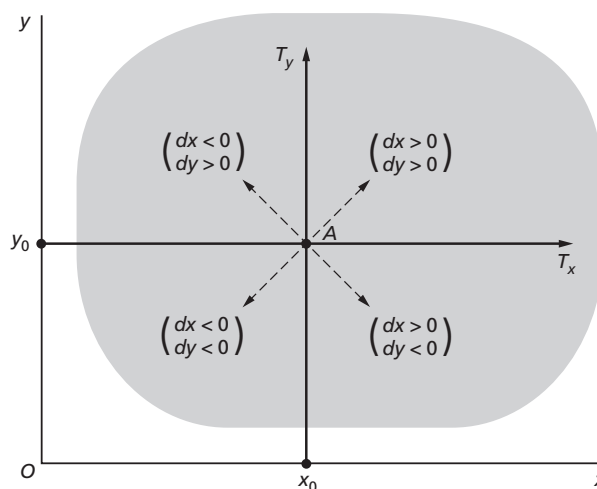
$$f_x = 24x^2 + 2y - 6x \quad f_y = 2x + 2y$$

$$f_{xx} = 48x - 6 \quad f_{yy} = 2 \quad f_{xy} = 2$$

A condição de primeira ordem exige a satisfação das equações simultâneas $f_x = 0$ e $f_y = 0$; isto é,

$$\begin{aligned} 24x^2 + 2y - 6x &= 0 \\ 2y + 2x &= 0 \end{aligned}$$

FIGURA 11.4



Condição	Máximo	Mínimo
Condição necessária de primeira ordem	$f_x = f_y = 0$	$f_x = f_y = 0$
Condição suficiente de segunda ordem [†]	$f_{xx}, f_{yy} < 0$ e $f_{xx} f_{yy} > f_{xy}^2$	$f_{xx}, f_{yy} > 0$ e $f_{xx} f_{yy} > f_{xy}^2$

[†] Aplicável somente após a condição necessária de primeira ordem ter sido satisfeita.

TABELA 11.1
Condições para
Extremo
Relativo:
 $z = f(x, y)$

A segunda equação implica que $y = -x$ e, quando essa informação é substituída na primeira equação, obtemos $24x^2 - 8x = 0$, que resulta no par de soluções

$$x_1^* = 0 \quad [\text{implicando } y_1^* = -x_1^* = 0]$$

$$x_2^* = \frac{1}{3} \quad [\text{implicando } y_2^* = -\frac{1}{3}]$$

Para aplicar a condição de segunda ordem, notamos que, quando

$$x_1^* = y_1^* = 0$$

f_{xx} resulta em -6 , enquanto f_{yy} é 2 , de modo que $f_{xx} f_{yy}$ é negativo e é necessariamente menor que um valor ao quadrado f_{xy}^2 . Isso não cumpre a condição de segunda ordem. O fato de f_{xx} e f_{yy} terem sinais opostos sugere, é claro, que a superfície em questão se curvará para cima em uma direção, mas para baixo em outra, dando origem, portanto, a um ponto de sela.

E a outra solução? Quando calculada em $x_2^* = \frac{1}{3}$, encontramos $f_{xx} = 10$, que, juntamente com o fato de que $f_{yy} = f_{xy} = 2$, cumpre todas as três partes da condição suficiente de segunda ordem para um mínimo. Portanto, fazendo $x = \frac{1}{3}$ e $y = -\frac{1}{3}$ na função dada, podemos obter, como um mínimo de z , o valor $z^* = \frac{23}{27}$. Assim, no presente exemplo, existe apenas um extremo relativo (um mínimo), que pode ser representado pela tripla ordenada

$$(x^*, y^*, z^*) = \left(\frac{1}{3}, -\frac{1}{3}, \frac{23}{27} \right)$$

EXEMPLO 5

Calcule o valor (ou valores) extremo de $z = x + 2ey - e^x - e^{2y}$. As derivadas relevantes dessa função são

$$\begin{aligned} f_x &= 1 - e^x & f_y &= 2e - 2e^{2y} \\ f_{xx} &= -e^x & f_{yy} &= -4e^{2y} & f_{xy} &= 0 \end{aligned}$$

Para satisfazer a condição necessária, devemos ter

$$\begin{aligned} 1 - e^x &= 0 \\ 2e - 2e^{2y} &= 0 \end{aligned}$$

que tem somente uma solução, a saber $x^* = 0$ e $y^* = \frac{1}{2}$. Para averiguar o status do valor de z correspondente a essa solução (o valor estacionário), calculamos as derivadas de segunda ordem em $x = 0$ e $y = \frac{1}{2}$ e constatamos que $f_{xx} = -1$, $f_{yy} = -4e$ e $f_{xy} = 0$. Visto que f_{xx} e f_{yy} são ambas negativas e, já que, além disso, $(-1)(-4e) > 0$, podemos concluir que o valor z em questão, a saber,

$$z^* = 0 + e - e^0 - e^1 = -1$$

é um valor máximo da função. Esse ponto máximo na superfície dada pode ser denotado pela tripla ordenada $(x^*, y^*, z^*) = (0, \frac{1}{2}, -1)$.

Observe, mais uma vez, que, para calcular as derivadas parciais segundas em x^* e y^* , deve-se fazer a diferenciação em primeiro lugar e, então, como etapa final, os valores específicos de x^* e y^* devem ser substituídos nas derivadas.

EXERCÍCIO 11.2

Use a Tabela 11.1 para calcular o valor (ou valores) extremo de cada uma das quatro funções seguintes, e determinar se são máximos ou mínimos:

1. $z = x^2 + xy + 2y^2 + 3$
2. $z = -x^2 - y^2 + 6x + 2y$
3. $z = ax^2 + by^2 + c$; considere cada um dos três subcasos:
 - (a) $a > 0, b > 0$
 - (b) $a < 0, b < 0$
 - (c) a e b com sinais opostos
4. $z = e^{2x} - 2x + 2y^2 + 3$
5. Considere a função $z = (x - 2)^4 + (y - 3)^4$.
 - (a) Estabeleça, por raciocínio intuitivo, que z atinge um mínimo ($z^* = 0$) em $x^* = 2$ e $y^* = 3$.
 - (b) A condição necessária de primeira ordem na Tabela 11.1 foi satisfeita?
 - (c) A condição suficiente de segunda ordem na Tabela 11.1 foi satisfeita?
 - (d) Calcule o valor de d^2z . Ele satisfaz a condição necessária de segunda ordem para um mínimo (11.9)?

11.3 Formas quadráticas – uma excursão

A expressão para d^2z na última linha de (11.6) exemplifica as denominadas *formas quadráticas*, para as quais existem critérios estabelecidos para determinar se seus sinais são sempre positivos, negativos, não-positivos ou não-negativos para valores arbitrários de dx e dy , não sendo ambos iguais a zero. Uma vez que a condição de segunda ordem para extremos depende diretamente do sinal de d^2z , esses critérios são de interesse direto.

Para começar, definimos uma *forma* como uma expressão polinomial na qual cada termo componente tem um grau uniforme. Quando examinamos polinômios pela primeira vez, nos limitamos ao caso de uma única variável: $a_0 + a_1x + \dots + a_nx^n$. Quando há mais variáveis envolvidas, cada termo de um polinômio pode conter ou uma variável ou diversas variáveis, cada uma elevada a uma potência inteira não-negativa, tal como $3x + 4x^2y^3 - 2yz$. No caso especial em que cada termo tem um grau uniforme – isto é, quando a soma de expoentes em cada termo for uniforme –, o polinômio é denominado uma *forma*. Por exemplo, $4x - 9y + z$ é uma *forma linear* com três variáveis, porque cada um de seus termos é de primeiro grau. Por outro lado, o polinômio $4x^2 - xy + 3y^2$, no qual cada termo é do segundo grau (soma dos expoentes inteiros = 2), constitui uma *forma quadrática* com duas variáveis. Também podemos encontrar formas quadráticas com três variáveis, tal como $x^2 + 2xy - yw + 7w^2$ ou, na verdade, com n variáveis.

Diferencial total de segunda ordem como uma forma quadrática

Se considerarmos as diferenciais dx e dy em (11.6) como variáveis e as derivadas parciais como coeficientes, isto é, se fizermos

$$\begin{aligned} u &\equiv dx & v &\equiv dy \\ a &\equiv f_{xx} & b &\equiv f_{yy} & h &\equiv f_{xy} [= f_{yx}] \end{aligned} \quad (11.10)$$

então a diferencial total de segunda ordem

$$d^2z = f_{xx} dx^2 + 2 f_{xy} dx dy + f_{yy} dy^2$$

pode ser facilmente identificada como uma forma quadrática q com duas variáveis u e v :

$$q = au^2 + 2huv + bv^2 \quad (11.6')$$

Note que, nessa forma quadrática, $dx \equiv u$ e $dy \equiv v$ fazem o papel de variáveis, ao passo que as derivadas parciais segundas são tratadas como constantes – o exato oposto da situação em que estávamos diferenciando dz para obter d^2z . A razão dessa inversão de papéis está na natureza diversa do problema com o qual estamos tratando agora. A condição suficiente de segunda ordem para extremos estipula que d^2z tem de ser definidamente positiva (para um mínimo) e definidamente negativa (para um máximo), independentemente dos valores que dx e dy possam assumir (contanto que não sejam ambos iguais a zero). Portanto, é óbvio que, no presente contexto, dx e dy devem ser consideradas como *variáveis*. As derivadas parciais segundas, por outro lado, assumirão valores específicos nos pontos que estamos examinando como possíveis pontos extremos e, assim, podem ser consideradas como *constantes*.

Então, a questão principal passa a ser: quais restrições devem ser impostas a a , b e h em (11.6'), enquanto permite-se que u e v assumam quaisquer valores, de modo a assegurar um sinal definido para q ?

Formas negativas definidas e positivas definidas

Por questão de terminologia, vamos observar que diz-se que uma forma quadrática q é

<i>Positiva definida</i>	se q for invariavelmente	positiva	(> 0)
<i>Positiva semidefinida</i>		não-negativa	(≥ 0)
<i>Negativa semidefinida</i>		não-positiva	(≤ 0)
<i>Negativa definida</i>		negativa	(< 0)

independente dos valores das variáveis na forma quadrática, não sendo todos iguais a zero. Se q mudar de sinal quando a variável assumir valores diferentes, por outro lado, diz-se que q é *indefinida*. Fica claro que os casos em que $q = d^2z$ é positiva definida ou negativa definida estão relacionados com as condições *suficientes* de segunda ordem para um mínimo e um máximo, respectivamente. Os casos em que a forma quadrática é *semidefinida*, por outro lado, estão relacionados com as condições *necessárias* de segunda ordem. Quando $q = d^2z$ for indefinida, temos o sintoma de um ponto de sela.

Teste com determinantes para condição de sinal definido

Um teste muito usado para a condição de sinal definido de q exige o exame dos sinais de certos determinantes. Acontece que esse teste é aplicável com maior facilidade para condição de positivo ou negativo definido (em oposição à condição de semidefinido); isto é, ele se aplica com maior facilidade em oposição a condições suficientes de segunda ordem (a condições necessárias). Aqui, nossa discussão ficará restrita apenas às condições suficientes.[†]

Para o caso de duas variáveis, é relativamente fácil obter condições envolvendo determinantes para sinal definido de q . Em primeiro lugar, vemos que os sinais do primeiro e do terceiro termos em (11.6') são independentes dos valores das variáveis u e v , porque essas variáveis estão elevadas ao quadrado. Por isso, é fácil especificar a condição para a definição de positiva

[†] Para uma discussão de um teste com determinantes para condições necessárias de segunda ordem, veja Chiang, Alpha C. *Elements of Dynamic Optimization*. Waveland Press Inc., 1992, p. 85–90.

definida ou negativa definida apenas desses termos, restringindo os sinais de a e b . O problema é o termo do meio. Mas, se pudermos converter todo o polinômio em uma expressão tal que as variáveis u e v apareçam somente em alguns termos elevados ao quadrado, será novamente possível tratar a condição de sinal definido de q .

O mecanismo que resolverá o problema é denominado completar o quadrado. Adicionando h^2v^2/a ao lado direito de (11.6'), e dele subtraindo a mesma quantidade, podemos reescrever a forma quadrática como se segue:

$$\begin{aligned} q &= au^2 + 2huv + \frac{h^2}{a}v^2 + bv^2 - \frac{h^2}{a}v^2 \\ &= a\left(u^2 + \frac{2h}{a}uv + \frac{h^2}{a^2}v^2\right) + \left(b - \frac{h^2}{a}\right)v^2 \\ &= a\left(u + \frac{h}{a}v\right)^2 + \frac{ab - h^2}{a}(v^2) \end{aligned}$$

Agora que as variáveis u e v aparecem apenas em termos elevados ao quadrado, podemos vincular o sinal de q inteiramente aos valores dos coeficientes a , b e h da seguinte maneira:

$$q \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ se e somente se } \begin{cases} a > 0 \\ a < 0 \end{cases} \text{ e } ab - h^2 > 0 \quad (11.11)$$

Agora que (1) $ab - h^2$ deve ser *positivo* em ambos os casos e (2) como um pré-requisito para a positividade de $ab - h^2$, o produto ab deve ser positivo (já que tem de ser maior que o termo ao quadrado h^2); então essa condição implica, automaticamente, que a e b devem assumir sinais algébricos idênticos.

A condição que acabamos de obter pode ser enunciada de modo mais sucinto utilizando determinantes. Observamos, em primeiro lugar, que a forma quadrática em (11.6') pode ser rearranjada no seguinte formato quadrado simétrico:

$$q = a(u^2) + h(uv) + h(vu) + b(v^2)$$

com os termos ao quadrado colocados na diagonal e com o termo $2huv$ dividido em duas partes iguais e colocado fora da diagonal. Agora, os coeficientes formam uma matriz simétrica, com a e b na diagonal principal e h fora da diagonal. Sob essa perspectiva, também é fácil ver a forma quadrática como a matriz 1×1 (um escalar) resultante da seguinte multiplicação matricial:

$$q = [u \ v] \begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Note que esse é um caso mais geral do produto matricial $x'Ax$ discutido na Seção 4.4, Exemplo 5. Naquele exemplo, com uma matriz denominada diagonal (uma matriz simétrica cujos elementos fora da diagonal são todos iguais a zero) como A , o produto $x'Ax$ representa uma soma ponderada de quadrados. Aqui, com qualquer matriz simétrica como A (permitindo que apareçam elementos diferentes de zero fora da diagonal), o produto $x'Ax$ é uma forma quadrática.

O determinante da matriz 2×2 de coeficientes, $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$ — que denominamos o *discriminante*

da forma quadrática q e que, portanto, denotaremos por $|D|$ —, dá a pista para o critério em (11.11), pois este último pode ser expresso alternativamente como:

$$q \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ se e somente se } \begin{cases} |a| > 0 \\ |a| < 0 \end{cases} \text{ e } \begin{vmatrix} a & h \\ h & b \end{vmatrix} > 0 \quad (11.11')$$

O determinante $|a| = a$ é simplesmente o *primeiro menor principal líder* de $|D|$. O determinante $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$, por outro lado, é o *segundo menor principal líder* de $|D|$. No caso presente, há somente dois menores principais líderes disponíveis, e seus sinais servirão para determinar a condição de q ser positiva definida ou negativa.

Quando (11.11') é traduzido, por meio de (11.10), para termos da diferencial total de segunda ordem d^2z , temos

$$d^2z \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ se e somente se } \begin{cases} f_{xx} > 0 \\ f_{xx} < 0 \end{cases} \text{ e } \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - f_{xy}^2 > 0$$

Recordando que a última desigualdade implica que f_{xx} e f_{yy} devem ter, como requisito, o *mesmo* sinal, vemos que isso é, exatamente, a condição suficiente de segunda ordem apresentada na Tabela 11.1.

Em geral, o discriminante de uma forma quadrática

$$q = au^2 + 2huv + bv^2$$

é o determinante simétrico $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$. No caso particular da forma quadrática

$$d^2z = f_{xx} dx^2 + 2 f_{xy} dx dy + f_{yy} dy^2$$

o discriminante é um determinante cujos elementos são as derivadas parciais de segunda ordem. Tal determinante é denominado um *determinante hessiano* (ou simplesmente um *hessiano*). No caso de duas variáveis, o hessiano é

$$|H| = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$$

o qual, em vista do teorema de Young ($f_{xy} = f_{yx}$), é simétrico – como um discriminante deve ser. É preciso distinguir cuidadosamente o determinante hessiano do determinante jacobiano discutido na Seção 7.6.

EXEMPLO 1 A forma quadrática $q = 5u^2 + 3uv + 2v^2$ é positiva definida ou negativa definida? O discriminante de q é $\begin{vmatrix} 5 & 1,5 \\ 1,5 & 2 \end{vmatrix}$, com menores principais líderes

$$5 > 0 \quad \text{e} \quad \begin{vmatrix} 5 & 1,5 \\ 1,5 & 2 \end{vmatrix} = 7,75 > 0$$

Portanto, q é positiva definida.

EXEMPLO 2 Dadas $f_{xx} = -2$, $f_{xy} = 1$ e $f_{yy} = -1$ em um certo ponto de uma função $z = f(x, y)$, d^2z tem um sinal definido naquele ponto independente dos valores de dx e dy ? O discriminante da forma quadrática d^2z é, neste caso, $\begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix}$, com menores principais líderes

$$-2 < 0 \quad \text{e} \quad \begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix} = 1 > 0$$

Assim, d^2z é negativa definida.

Formas quadráticas com três variáveis

Podem ser obtidas condições semelhantes para uma forma quadrática com *três* variáveis?

Uma forma quadrática com três variáveis u_1 , u_2 e u_3 pode ser representada de modo geral como

$$\begin{aligned} q(u_1, u_2, u_3) &= d_{11}(u_1^2) + d_{12}(u_1 u_2) + d_{13}(u_1 u_3) \\ &\quad + d_{21}(u_2 u_1) + d_{22}(u_2^2) + d_{23}(u_2 u_3) \\ &\quad + d_{31}(u_3 u_1) + d_{32}(u_3 u_2) + d_{33}(u_3^2) \\ &= \sum_{i=1}^3 \sum_{j=1}^3 d_{ij} u_i u_j \end{aligned} \quad (11.12)$$

onde a notação duplo- $\sum$ (duplo somatório) significa que é permitido que ambos os índices, i e j , assumam os valores 1, 2 e 3; e, por isso, a expressão do duplo somatório é equivalente ao arranjo 3×3 mostrado na Equação (11.12). A propósito, esse arranjo quadrado da forma quadrática deve ser sempre considerado como um arranjo simétrico, mesmo que tenhamos escrito o par de coeficientes (d_{12}, d_{21}) ou (d_{23}, d_{32}) como se os dois membros de cada par fossem diferentes. Pois, se acaso o termo na forma quadrática que envolve as variáveis u_1 e u_2 for, digamos, $12u_1 u_2$, podemos fazer $d_{12} = d_{21} = 6$, de modo que $d_{12}u_1 u_2 = d_{21}u_2 u_1$, e um procedimento similar pode ser aplicado para tornar simétricos os outros elementos fora da diagonal.

Na verdade, essa forma quadrática de três variáveis pode ser expressa, novamente, como um produto de três matrizes:

$$q(u_1, u_2, u_3) = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \equiv u' D u \quad (11.12')$$

Como no caso de duas variáveis, a primeira matriz (um vetor linha) e a terceira matriz (um vetor coluna) apenas relacionam as variáveis, e a do meio (D) é uma matriz simétrica de coeficientes da versão em arranjo quadrado da forma quadrática em (11.12). Desta vez, entretanto, pode ser formado um total de *três* menores principais líderes a partir de seu discriminante, a saber,

$$|D_1| \equiv d_{11} \quad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \quad |D_3| \equiv \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix}$$

onde $|D_i|$ denota o i -ésimo menor principal líder do discriminante $|D|$.[†] Acontece que as condições de positiva definida ou negativa definida podem novamente ser enunciadas em termos de certas restrições de sinais impostas a esses menores principais.

Pelo mecanismo agora já familiar de completar o quadrado, a forma quadrática em (11.12) pode ser convertida em uma expressão na qual as três variáveis aparecem somente como componentes de alguns termos ao quadrado. Especificamente, lembrando que $a_{12} = a_{21}$ etc., temos

[†] Até agora examinamos o i -ésimo menor principal líder $|D_i|$ como um subdeterminante formado pela retenção dos primeiros i elementos da diagonal principal de $|D|$. Entretanto, uma vez que a noção de um *menor* implica a *exclusão* de alguma coisa do determinante original, você talvez prefira considerar o i -ésimo menor principal líder alternativamente como um subdeterminante formado pela exclusão das últimas $(n - i)$ linhas e colunas de $|D|$.

$$q = d_{11} \left(u_1 + \frac{d_{12}}{d_{11}} u_2 + \frac{d_{13}}{d_{11}} u_3 \right)^2 + \frac{d_{11}d_{22} - d_{12}^2}{d_{11}} \left(u_2 + \frac{d_{11}d_{23} - d_{12}d_{13}}{d_{11}d_{22} - d_{12}^2} u_3 \right)^2 + \frac{d_{11}d_{22}d_{33} - d_{11}d_{23}^2 - d_{22}d_{13}^2 - d_{33}d_{12}^2 + 2d_{12}d_{13}d_{23}}{d_{11}d_{22} - d_{12}^2} (u_3)^2$$

Essa soma de quadrados será positiva (negativa) para quaisquer valores de u_1 , u_2 e u_3 , não sendo todos iguais a zero, se e somente se os coeficientes das três expressões elevadas ao quadrado forem todos positivos (negativos). Mas os três coeficientes (na ordem dada) podem ser expressos em termos dos três menores principais líderes como se segue:

$$|D_1| \quad \frac{|D_2|}{|D_1|} \quad \frac{|D_3|}{|D_2|}$$

Portanto, para a condição de *positiva* definida, a condição necessária e suficiente é tripla:

$$\begin{aligned} |D_1| &> 0 \\ |D_2| &> 0 \quad [\text{dado que } |D_1| > 0 \text{ antes}] \\ |D_3| &> 0 \quad [\text{dado que } |D_2| > 0 \text{ antes}] \end{aligned}$$

Em outras palavras, os três menores principais líderes devem ser todos positivos. Para a condição de *negativa* definida, por outro lado, a condição necessária e suficiente se torna:

$$\begin{aligned} |D_1| &< 0 \\ |D_2| &> 0 \quad [\text{dado que } |D_1| < 0 \text{ antes}] \\ |D_3| &< 0 \quad [\text{dado que } |D_2| > 0 \text{ antes}] \end{aligned}$$

Isto é, os três menores principais líderes devem alternar sinais da maneira especificada.

EXEMPLO 3 Determine se $q = u_1^2 + 6u_2^2 + 3u_3^2 - 2u_1u_2 - 4u_2u_3$ é positiva definida ou negativa definida. O discriminante de q é

$$\begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix}$$

com menores principais líderes como se segue:

$$1 > 0 \quad \begin{vmatrix} 1 & -1 \\ -1 & 6 \end{vmatrix} = 5 > 0 \quad \text{e} \quad \begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix} = 11 > 0$$

Portanto, a forma quadrática é positiva definida.

EXEMPLO 4 Determine se $q = 2u^2 + 3v^2 - w^2 + 6uv - 8uw - 2vw$ é positiva definida ou negativa definida. O discriminante pode ser escrito como

$$\begin{vmatrix} 2 & 3 & -4 \\ 3 & 3 & -1 \\ -4 & -1 & -1 \end{vmatrix}, \text{ e constatamos que seu primeiro menor principal}$$

$$\text{líder é } 2 > 0, \text{ mas o segundo menor principal líder é } \begin{vmatrix} 2 & 3 \\ 3 & 3 \end{vmatrix} = -3 < 0.$$

Isso viola a condição tanto para positiva definida quanto para negativa definida; assim, q não é nem positiva definida nem negativa definida.

Formas quadráticas com n variáveis

Como uma extensão do resultado precedente para o caso de n variáveis, vamos declarar, sem provar, que, para a forma quadrática

$$q(u_1, u_2, \dots, u_n) = \sum_{i=1}^n \sum_{j=1}^n d_{ij} u_i u_j \quad [\text{onde } d_{ij} = d_{ji}]$$

$$= \underset{(1 \times n)}{u'} \underset{(n \times n)}{D} \underset{(n \times 1)}{u} \quad [\text{veja (11.12')}]$$

a condição necessária e suficiente para ser *positiva definida* é que os menores principais líderes de $|D|$, a saber,

$$|D_1| \equiv d_{11} \quad |D_2| \equiv \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \quad \dots \quad |D_n| \equiv \begin{vmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{vmatrix}$$

sejam todos positivos. A condição necessária e suficiente correspondente para *negativa definida* é que os menores principais líderes alternem sinais como se segue:

$$|D_1| < 0 \quad |D_2| > 0 \quad |D_3| < 0 \quad (\text{etc.})$$

de modo que todos os de número *ímpar* sejam negativos e todos os de número *par* sejam positivos. O n -ésimo menor principal líder, $|D_n| = |D|$, deve ser positivo se n for par, mas negativo se n for ímpar, o que pode ser expresso sucintamente pela desigualdade $(-1)^n |D_n| > 0$.

Teste da raiz característica para condição de sinal definido

À parte o teste com determinantes precedentes para a condição de sinal definido de uma forma quadrática $u'Du$, há um teste alternativo que utiliza o conceito das denominadas raízes características da matriz D . Esse conceito surge em um problema da seguinte natureza. Dada uma matriz $n \times n$, D , podemos achar um escalar r e um vetor $n \times 1$ $x \neq 0$, de modo que a equação matricial

$$\underset{(n \times n)}{D} \underset{(n \times 1)}{x} = r \underset{(n \times 1)}{x} \quad (11.13)$$

seja satisfeita? Se for possível, o escalar r é denominado uma *raiz característica* da matriz D e x é denominado um *vetor característico* daquela matriz.[†]

A equação matricial $Dx = rx$ pode ser reescrita como $Dx - rIx = 0$, ou

$$(D - rI)x = 0 \quad \text{onde } 0 \text{ é } n \times 1 \quad (11.13')$$

Isso, é claro, representa um sistema de n equações lineares homogêneas. Visto que queremos uma solução não-trivial para x , a matriz de coeficientes $(D - rI)$ – denominada a *matriz característica* de D – tem de ser singular. Em outras palavras, seu determinante tem de ser nulo:

$$|D - rI| = \begin{vmatrix} d_{11} - r & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} - r & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} - r \end{vmatrix} = 0 \quad (11.14)$$

[†] Raízes características também são conhecidas pelos nomes alternativos de *raízes latentes* ou *autovalores*. Vetores característicos também são denominados *autovetores*.

A equação (11.14) é denominada a *equação característica* da matriz D . Uma vez que, com expansão de Laplace, o determinante $|D - rI|$ resultará em um polinômio de grau n na variável r , (11.14) é, na verdade, uma equação polinomial de n -ésimo grau. Assim, haverá um total de n raízes $(r_1, \dots, r_n)$, e cada uma delas se qualifica como uma raiz característica. Se D for simétrico, como no contexto da forma quadrática, as raízes características sempre resultarão em números reais, que podem ter qualquer sinal algébrico ou ser nulos.

Considerando que todos esses valores de r anularão o determinante $|D - rI|$, a substituição de qualquer um deles (digamos, r_i) no sistema de equações (11.13') produzirá um vetor correspondente $x|_{r=r_i}$. Mais precisamente, por ser homogêneo, o sistema dará um número infinito de vetores que correspondem à raiz r_i . Contudo, aplicaremos um processo de *normalização* (que será explicado no Exemplo 5) e selecionaremos um elemento particular daquele conjunto infinito como o vetor característico correspondente a r_i ; esse vetor será denotado por v_i . Com um total de n raízes características, deve haver um total de n desses vetores característicos correspondentes.

EXEMPLO 5

Calcule as raízes e vetores característicos da matriz $\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$. Substituindo D pela matriz dada em (11.14), obtemos a equação

$$\begin{vmatrix} 2-r & 2 \\ 2 & -1-r \end{vmatrix} = r^2 - r - 6 = 0$$

cuja raízes são $r_1 = 3$ e $r_2 = -2$. Quando a primeira raiz é usada, a equação matricial (11.13') toma a forma de

$$\begin{bmatrix} 2-3 & 2 \\ 2 & -1-3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Como as duas linhas da matriz de coeficientes são linearmente dependentes, como seria de se esperar em vista de (11.14), há um número infinito de soluções que pode ser expresso pela equação $x_1 = 2x_2$. Para forçar uma solução única, *normalizamos* a solução impondo a restrição $x_1^2 + x_2^2 = 1$.[†] Então, visto que

$$x_1^2 + x_2^2 = (2x_2)^2 + x_2^2 = 5x_2^2 = 1$$

podemos obter (tomando a raiz quadrada positiva) $x_2 = 1/\sqrt{5}$, e também $x_1 = 2x_2 = 2/\sqrt{5}$. Assim, o primeiro vetor característico é

$$v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

De modo semelhante, usando a segunda raiz $r_2 = -2$ em (11.13'), obtemos a equação

$$\begin{bmatrix} 2-(2) & 2 \\ 2 & -1-(-2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

cuj solução é $x_1 = -\frac{1}{2}x_2$. Com a normalização, obtemos

$$x_1^2 + x_2^2 = \left(-\frac{1}{2}x_2\right)^2 + x_2^2 = \frac{5}{4}x_2^2 = 1$$

que resulta em $x_2 = 2/\sqrt{5}$ e $x_1 = -1/\sqrt{5}$. Assim, o segundo vetor característico é

$$v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

[†] De modo mais geral, para o caso de n variáveis, a restrição é $\sum_{i=1}^n x_i^2 = 1$.

O conjunto de vetores característicos obtidos dessa maneira possui duas propriedades importantes. A primeira é que o produto escalar $v_i' v_i$ ($i = 1, 2, \dots, n$) deve ser igual à unidade, uma vez que

$$v_i' v_i = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n x_i^2 = 1 \quad [\text{por normalização}]$$

A segunda é que o produto escalar $v_i' v_j$ (onde $i \neq j$) sempre pode ser tomado como zero.[†] Portanto, em suma, podemos escrever que

$$v_i' v_i = 1 \quad \text{e} \quad v_i' v_j = 0 \quad (i \neq j) \quad (11.15)$$

Essas propriedades se revelarão úteis mais tarde (veja o Exemplo 6). Por questão de terminologia, quando dois vetores dão um produto escalar de valor zero, diz-se que eles são *ortogonais* (perpendiculares) um ao outro.[‡] Portanto, cada par de vetores característicos da matriz D deve ser ortogonal. A outra propriedade, $v_i' v_i = 1$, indica a normalização. Juntas, essas duas propriedades são responsáveis pelo fato de dizermos que os vetores característicos ($v_1, \dots, v_n$) são um conjunto de vetores *ortonormais*. Tente verificar a ortonormalidade dos dois vetores característicos encontrados no Exemplo 5.

Agora estamos prontos para explicar como as raízes características e os vetores característicos da matriz D podem ser úteis para determinar a condição de sinal definido da forma quadrática $u' Du$. Em essência, a idéia é, novamente, transformar $u' Du$ (que envolve não somente termos ao quadrado $u_1^2, \dots, u_n^2$, mas também termos de produto cruzado como $u_1 u_2$ e $u_2 u_3$) em uma forma que contenha apenas termos ao quadrado. Assim, a abordagem é semelhante à do processo de completar o quadrado usado anteriormente para derivar o teste com determinantes. Contudo, no caso presente, a transformação possui uma característica adicional: cada termo ao quadrado tem como seu coeficiente uma das raízes características, de modo que os sinais das n raízes fornecerão informação suficiente para determinar a condição de sinal definido da forma quadrática.

A transformação que fará isso é a seguinte. Vamos deixar que os vetores característicos $v_1, \dots, v_n$ constituam as *colunas* de uma matriz T :

$$T_{(n \times n)} = [v_1 \quad v_2 \quad \cdots \quad v_n]$$

e então vamos aplicar a transformação $u = T y$ à forma quadrática $u' Du$:

[†] Para demonstrar isso, notamos que, por (11.13), podemos escrever $Dv_j = r_j v_j$, e $Dv_i = r_i v_i$. Pré-multiplicando ambos os lados de cada uma dessas equações por um vetor linha adequado, temos

$$v_i' Dv_j = v_i' r_j v_j = r_j v_i' v_j \quad [r_j \text{ é um escalar}]$$

$$v_j' Dv_i = v_j' r_i v_i = r_i v_j' v_i = r_i v_i' v_j \quad [v_i' v_j = v_j' v_i]$$

Visto que $v_i' Dv_j$ e $v_j' Dv_i$ são ambos 1×1 e, uma vez que são transposições um do outro (lembre-se que $D' = D$ porque D é simétrico), eles devem representar o mesmo escalar. Deduz-se que as expressões na extrema direita dessas duas equações são iguais; portanto, por subtração, temos

$$(r_j - r_i) v_i' v_j = 0$$

Agora, se $r_j \neq r_i$ (raízes distintas), então $v_i' v_j$ tem de ser zero para que a equação seja válida, e isso ratifica o que declaramos. Se, além do mais, $r_j = r_i$ (raízes repetidas), sempre será possível, afinal, achar dois vetores normalizados linearmente independentes que satisfaçam $v_i' v_j = 0$. Assim, podemos declarar, em geral, que $v_i' v_j = 0$, sempre que $i \neq j$.

[‡] À guisa de uma explicação simples, imagine os dois vetores unidade de um espaço bidimensional, $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ e $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Esses vetores

estão, respectivamente, sobre os dois eixos, e, por isso, são perpendiculares. Ao mesmo tempo, constatamos realmente que $e_1' e_2 = e_2' e_1 = 0$. Veja também o Exercício 4.3–4.

$$u'Du = (T'y)'D(Ty) = y'T'DTy \quad [\text{por (4.11)}]$$

$$= y'Ry \quad \text{onde} \quad R \equiv T'DT$$

O resultado é que a forma quadrática original nas variáveis u_i transforma-se agora em uma outra forma quadrática nas variáveis y_i . Visto que as variáveis u_i e as variáveis y_i assumem a mesma faixa de valores, a transformação não afeta a condição de sinal definido da forma quadrática. Assim, agora podemos perfeitamente considerar o sinal da forma quadrática $y'Ry$ em seu lugar. O que torna esta última forma quadrática intrigante é que a matriz R se revelará uma matriz diagonal, sendo que as raízes $r_1, \dots, r_n$ da matriz D aparecem ao longo de sua diagonal e em todos os outros lugares aparecem zeros, de modo que temos, de fato,

$$u'Du = y'Ry = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (11.16)$$

$$= r_1 y_1^2 + r_2 y_2^2 + \cdots + r_n y_n^2$$

que é uma expressão que envolve apenas termos ao quadrado. A transformação $R \equiv T'DT$ nos fornece, portanto, um procedimento para *diagonalizar* a matriz simétrica D para uma matriz diagonal especial R .

EXEMPLO 6

Verifique se a matriz $\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$ dada no Exemplo 5 pode ser diagonalizada para a matriz

$\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$. Com base nos vetores característicos encontrados no Exemplo 5, a matriz de transformação T deve ser

$$T = [v_1 \quad v_2] = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

Assim, podemos escrever

$$R \equiv T'DT = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$$

que verifica devidamente o processo de diagonalização.

Para provar o resultado da diagonalização em (11.16) vamos escrever (parcialmente) a matriz R da seguinte maneira:

$$R \equiv T'DT = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix} D \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

Podemos verificar com facilidade que $D[v_1 \ v_2 \ \cdots \ v_n]$ pode ser reescrita como $[Dv_1 \ Dv_2 \ \cdots \ Dv_n]$. Além disso, por (11.13), podemos ainda reescrever essa expressão como $[r_1 v_1 \ r_2 v_2 \ \cdots \ r_n v_n]$. Portanto, verificamos que

$$R = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix} \begin{bmatrix} r_1 v_1 & r_2 v_2 & \cdots & r_n v_n \end{bmatrix} = \begin{bmatrix} r_1 v'_1 v_1 & r_2 v'_1 v_2 & \cdots & r_n v'_1 v_n \\ r_1 v'_2 v_1 & r_2 v'_2 v_2 & \cdots & r_n v'_2 v_n \\ \cdots & \cdots & \cdots & \cdots \\ r_1 v'_n v_1 & r_2 v'_n v_2 & \cdots & r_n v'_n v_n \end{bmatrix}$$

$$= \begin{bmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & r_n \end{bmatrix} \quad [\text{por (11.15)}]$$

que é exatamente o que pretendíamos mostrar.

Em vista do resultado em (11.16), podemos enunciar formalmente o teste da raiz característica para a condição de sinal definido da forma quadrática da seguinte maneira:

1. $q = u'Du$ é positiva (negativa) definida se e somente se *cada* raiz característica de D for positiva (negativa).
2. $q = u'Du$ é positiva (negativa) semidefinida se e somente se *todas* as raízes características de D forem não-negativas (não-positivas).
3. $q = u'Du$ é indefinida se e somente se algumas das raízes características de D forem positivas e algumas negativas.

Note que, ao aplicar esse teste, bastam as raízes características; não precisamos dos vetores característicos a menos que queiramos encontrar a matriz de transformação T . Note, também, que, esse teste, diferentemente do teste com determinantes esboçado antes, nos permite verificar as condições necessárias de segunda ordem (parte 2 do teste) simultaneamente com as condições suficientes (parte 1 do teste). Porém, ele tem uma desvantagem. Quando a dimensão da matriz D for grande, a equação polinomial (11.14) pode não ser resolvida com facilidade para as raízes características exigidas para o teste. Nesses casos, o teste com determinantes ainda pode ser preferível.

EXERCÍCIO 11.3

1. Por multiplicação matricial direta, expresse cada um dos seguintes produtos de matrizes como uma forma quadrática:

$$(a) \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$(c) \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$(b) \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$(d) \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}$$

2. Nos Problemas 1b e c, as matrizes de coeficientes não são simétricas em relação à diagonal principal. Verifique se, tomando a média dos elementos fora da diagonal e assim os convertendo respectivamente para $\begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix}$ e $\begin{bmatrix} 5 & 3 \\ 3 & 0 \end{bmatrix}$, obteremos as mesmas formas quadráticas de antes.
3. Com base em suas matrizes de coeficientes (as versões *simétricas*), determine, pelo teste com determinantes, se as formas quadráticas nos Problemas 1a, b e c são positivas definidas ou negativas definidas.
4. Expresse cada uma das seguintes formas quadráticas como um produto de matrizes envolvendo uma matriz de coeficientes *simétrica*.

$$(a) q = 3u^2 - 4uv + 7v^2$$

$$(d) q = 6xy - 5y^2 - 2x^2$$

$$(b) q = u^2 + 7uv + 3v^2$$

$$(e) q = 3u_1^2 - 2u_1u_2 + 4u_1u_3 + 5u_2^2 + 4u_3^2 - 2u_2u_3$$

$$(c) q = 8uv - u^2 - 31v^2$$

$$(f) q = -u^2 + 4uv - 6uw - 4v^2 - 7w^2$$

5. A partir dos determinantes obtidos das matrizes de coeficientes simétricas do Problema 4, averigüe, pelo teste com determinantes, quais das formas quadráticas são positivas definidas e quais são negativas definidas.
6. Encontre as raízes características de cada uma das seguintes matrizes:

$$(a) D = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix} \quad (b) E = \begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix} \quad (c) F = \begin{bmatrix} 5 & 3 \\ 3 & 0 \end{bmatrix}$$

O que você pode concluir sobre os sinais das formas quadráticas $u'Du$, $u'Eu$ e $u'Fu$? (Verifique se seus resultados estão de acordo com o Problema 3.)

7. Encontre os vetores característicos da matriz $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.
8. Dada a forma quadrática $u'Du$, onde D é 2×2 , a equação característica de D pode ser escrita como

$$\begin{vmatrix} d_{11}-r & d_{12} \\ d_{21} & d_{22}-r \end{vmatrix} = 0 \quad (d_{12} = d_{21})$$

Expanda o determinante; expresse as raízes dessa equação utilizando a fórmula da forma quadrática; e deduza o seguinte:

- (a) Nenhum número imaginário (um número envolvendo $\sqrt{-1}$) pode ocorrer em r_1 e r_2 .
- (b) Para ter raízes repetidas, a matriz D deve estar na forma de $\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$.
- (c) Para ser positiva semidefinida ou negativa semidefinida, o discriminante da forma quadrática pode se anular, isto é, $|D| = 0$ é possível.

11.4 Funções objetivo com mais de duas variáveis

Quando aparecem $n > 2$ variáveis de escolha em uma função objetivo, já não é mais possível representar a função gratificante, se bem que ainda possamos falar de uma *hipersuperfície* em um espaço de $(n + 1)$ dimensões. Em tal hipersuperfície (não representável gratificante), novamente podem existir análogos de picos de domos e fundos de tigela de $(n + 1)$ dimensões. Como nós os identificamos?

Condição de primeira ordem para extremo

Vamos considerar especificamente uma função com três variáveis de escolha

$$z = f(x_1, x_2, x_3)$$

cujas derivadas parciais primeiras são f_1 , f_2 e f_3 e cujas derivadas parciais segundas são f_{ij} ($\equiv \partial^2 z / \partial x_i \partial x_j$), com $i, j = 1, 2, 3$. Em virtude do teorema de Young, temos $f_{ij} = f_{ji}$.

Nossa discussão anterior sugere que, para ter um máximo ou um mínimo de z , é necessário que $dz = 0$ para valores arbitrários de dx_1 , dx_2 e dx_3 , nem todos iguais a zero. Visto que o valor de dz agora é

$$dz = f_1 dx_1 + f_2 dx_2 + f_3 dx_3 \quad (11.17)$$

e uma vez que dx_1 , dx_2 e dx_3 são variações arbitrárias nas variáveis independentes, nem todas iguais a zero, o único modo de garantir que dz seja zero é ter $f_1 = f_2 = f_3 = 0$. Assim, mais uma vez, a condi-

ção necessária para extremo é que todas as derivadas parciais de primeira ordem sejam zero, do mesmo modo que para o caso de duas variáveis.[†]

Condição de segunda ordem

A satisfação da condição de primeira ordem assinala certos valores de z como e valores estacionários da função objetivo. Se constataremos que, em um valor estacionário de z , d^2z é positiva definida, isso será suficiente para estabelecer aquele valor de z como um mínimo. Analogamente, a condição de negativa definida de d^2z é uma condição suficiente para que o valor estacionário seja um máximo. Isso levanta a questão de como expressar d^2z quando há três variáveis na função e de como determinar sua condição de positiva definida ou negativa definida.

A expressão para d^2z pode ser obtida pela deferenciação de dz em (11.17). Em tal processo, assim como em (11.6), devemos tratar as derivadas f_i como variáveis e as diferenciais dx_i como constantes. Desse modo, temos,

$$\begin{aligned} d^2z &= d(dz) = \frac{\partial(dz)}{\partial x_1} dx_1 + \frac{\partial(dz)}{\partial x_2} dx_2 + \frac{\partial(dz)}{\partial x_3} dx_3 \\ &= \frac{\partial}{\partial x_1} (f_1 dx_1 + f_2 dx_2 + f_3 dx_3) dx_1 \\ &\quad + \frac{\partial}{\partial x_2} (f_1 dx_1 + f_2 dx_2 + f_3 dx_3) dx_2 \\ &\quad + \frac{\partial}{\partial x_3} (f_1 dx_1 + f_2 dx_2 + f_3 dx_3) dx_3 \\ &= f_{11} dx_1^2 + f_{12} dx_1 dx_2 + f_{13} dx_1 dx_3 \\ &\quad + f_{21} dx_2 dx_1 + f_{22} dx_2^2 + f_{23} dx_2 dx_3 \\ &\quad + f_{31} dx_3 dx_1 + f_{32} dx_3 dx_2 + f_{33} dx_3^2 \end{aligned} \quad (11.18)$$

que é uma forma quadrática semelhante a (11.12). Conseqüentemente, os critérios para condição de positiva definida e negativa definida que aprendemos antes podem ser diretamente aplicados aqui.

Ao determinar a condição de positiva definida ou negativa definida de d^2z , devemos, novamente, como já fizemos em (11.6'), considerar dx_i como variáveis que podem assumir quaisquer valores (embora nem todos iguais a zero) e, ao mesmo tempo, considerar as derivadas f_{ij} como coeficientes sobre os quais impor certas restrições. Os coeficientes em (11.18) dão origem ao determinante hessiano simétrico

$$|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix}$$

cujos menores principais líderes podem ser denotados por

[†] Como um caso especial, note que, se estivermos trabalhando com uma função $z = f(x_1, x_2, x_3)$ implicitamente definida por uma equação $F(z, x_1, x_2, x_3) = 0$, onde

$$f_i \equiv \frac{\partial z}{\partial x_i} = \frac{-\partial F / \partial x_i}{\partial F / \partial z} \quad (i = 1, 2, 3)$$

então a condição de primeira ordem $f_1 = f_2 = f_3 = 0$ equivalerá à condição

$$\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = \frac{\partial F}{\partial x_3} = 0$$

já que o valor do denominador $\partial F / \partial z \neq 0$ não faz nenhuma diferença.

$$|H_1| = f_{11} \quad |H_2| = \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} \quad |H_3| = |H|$$

Assim, com base no critério com determinantes para a condição de positiva definida ou negativa definida, podemos enunciar a condição suficiente de segunda ordem para um extremo de z da seguinte maneira:

$$z^* \text{ é um } \begin{cases} \text{máximo} \\ \text{mínimo} \end{cases}$$

$$\text{se } \begin{cases} |H_1| < 0; & |H_2| > 0; & |H_3| < 0 \quad (d^2z \text{ negativa definida}) \\ |H_1| > 0; & |H_2| > 0; & |H_3| > 0 \quad (d^2z \text{ positiva definida}) \end{cases} \quad (11.19)$$

Ao usar essa condição, devemos avaliar todos os menores principais líderes no ponto estacionário onde $f_1 = f_2 = f_3 = 0$.

É claro que também podemos aplicar o teste da raiz característica e associar a condição de positiva definida (negativa definida) de d^2z com a positividade (negatividade) de todas as raízes

características da *matriz hessiana* $\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$. De fato, em vez de dizer que a diferencial to-

tal de segunda ordem d^2z é positiva (negativa) definida, também é aceitável afirmar que a matriz hessiana H (que não deve ser confundida com o determinante hessiano $|H|$) é positiva (negativa) definida. Ao fazer uso disso, entretanto, note que a condição de sinal definido de H se refere ao sinal da forma quadrática d^2z com a qual H está associado e *não* com os sinais dos elementos de H em si.

EXEMPLO 1 Calcule o valor (ou os valores) extremo de

$$z = 2x_1^2 + x_1x_2 + 4x_2^2 + x_1x_3 + x_3^2 + 2$$

A condição de primeira ordem para extremos envolve a satisfação simultânea das três equações seguintes:

$$\begin{aligned} (f_1 =) & 4x_1 + x_2 + x_3 = 0 \\ (f_2 =) & x_1 + 8x_2 = 0 \\ (f_3 =) & x_1 + 2x_3 = 0 \end{aligned}$$

Como esse é um sistema homogêneo de equações lineares, no qual todas as três equações são independentes (o determinante da matriz de coeficientes não é nulo), existe somente a solução única $x_1^* = x_2^* = x_3^* = 0$. Isso significa que há apenas um valor estacionário: $z^* = 2$.

O determinante hessiano dessa função é

$$|H| = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} = \begin{vmatrix} 4 & 1 & 1 \\ 1 & 8 & 0 \\ 1 & 0 & 2 \end{vmatrix}$$

cujos menores principais líderes são todos positivos:

$$|H_1| = 4 \quad |H_2| = 31 \quad |H_3| = 54$$

Assim, podemos concluir, por (11.9), que $z^* = 2$ é um mínimo.

EXEMPLO 2 Calcule o valor (ou valores) extremo de

$$z = -x_1^3 + 3x_1x_3 + 2x_2 - x_2^2 - 3x_3^2$$

As derivadas parciais primeiras são

$$f_1 = -3x_1^2 + 3x_3 \quad f_2 = 2 - 2x_2 \quad f_3 = 3x_1 - 6x_3$$

Igualando todas as f_i a zero, obtemos três equações simultâneas, uma não-linear e duas lineares:

$$\begin{aligned} -3x_1^2 + 3x_3 &= 0 \\ -2x_2 &= -2 \\ 3x_1 - 6x_3 &= 0 \end{aligned}$$

Visto que a segunda equação resulta em $x_2^* = 1$ e a terceira equação implica $x_1^* = 2x_3^*$, substituindo-as na primeira equação temos duas soluções:

$$(x_1^*, x_2^*, x_3^*) = \begin{cases} (0, 1, 0), & \text{implicando } z^* = 1 \\ \left(\frac{1}{2}, 1, \frac{1}{4}\right), & \text{implicando } z^* = \frac{17}{16} \end{cases}$$

As derivadas parciais de segunda ordem, arranjadas adequadamente, nos dão o Hessiano

$$|H| = \begin{vmatrix} -6x_1 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{vmatrix}$$

no qual o primeiro elemento ($-6x_1$) se reduz a 0 sob a primeira solução (com $x_1^* = 0$) e a -3 sob a segunda (com $x_1^* = \frac{1}{2}$). Fica imediatamente óbvio que a primeira solução não satisfaz a condição suficiente de segunda ordem, já que $|H_1| = 0$. Entretanto, podemos recorrer ao teste da raiz característica para mais informações. Com essa finalidade, aplicamos a equação característica (11.14). Posto que a forma quadrática que está sendo testada é d^2z , cujo discriminante é o determinante hessiano, devemos, é claro, substituir os elementos do hessiano pelos elementos d_{ij} naquela equação. Por conseguinte, a equação característica é (para a primeira solução)

$$\begin{vmatrix} -r & 0 & 3 \\ 0 & -2-r & 0 \\ 3 & 0 & -6-r \end{vmatrix} = 0$$

a qual, por expansão, torna-se a equação cúbica

$$r^3 + 8r^2 + 3r - 18 = 0$$

Usando o Teorema I da Seção 3.3, podemos achar uma raiz inteira -2 . Assim, a função cúbica deve ser divisível por $(r + 2)$, e podemos fatorar a função cúbica e reescrever a equação precedente como

$$(r + 2)(r^2 + 6r - 9) = 0$$

O termo $(r + 2)$ deixa claro que uma das raízes características é $r_1 = -2$. As duas outras raízes podem ser achadas pela aplicação da fórmula quadrática ao outro termo; elas são $r_2 = -3 + \frac{1}{2}\sqrt{72}$, e $r_3 = -3 - \frac{1}{2}\sqrt{72}$. Considerando que r_1 e r_3 são negativas, mas r_2 é positiva, a forma quadrática d^2z é indefinida e, portanto, viola as condições necessárias de segunda ordem tanto para um máximo quanto para um mínimo z . Assim, a primeira solução ($z^* = 1$) não é, de modo algum, um extremo.

Quanto à segunda solução, a situação é mais simples. Visto que os sinais dos menores principais líderes

$$|H_1| = -3 \quad |H_2| = 6 \quad \text{e} \quad |H_3| = -18$$

têm sinais devidamente alternados, o teste com determinantes é conclusivo. De acordo com (11.19), a solução $z^* = \frac{17}{16}$ é um máximo.

O caso de n variáveis

Quando há n variáveis de escolha, a função objetivo pode ser expressa como

$$z = f(x_1, x_2, \dots, x_n)$$

Então, a diferencial total será

$$dz = f_1 dx_1 + f_2 dx_2 + \dots + f_n dx_n$$

de modo que a condição necessária para extremos ($dz = 0$ para dx_i arbitrários, nem todos iguais a zero) significa que todas as n derivadas parciais de primeira ordem têm de ser iguais a zero.

A diferencial de segunda ordem d^2z novamente será uma forma quadrática, analogamente derivável para (11.18), e que pode ser expressa por um arranjo $n \times n$. Os coeficientes desse arranjo, adequadamente arrumados, agora nos darão o hessiano (simétrico)

$$|H| = \begin{vmatrix} f_{11} & f_{12} & \dots & f_{1n} \\ f_{21} & f_{22} & \dots & f_{2n} \\ \dots & \dots & \dots & \dots \\ f_{n1} & f_{n2} & \dots & f_{nn} \end{vmatrix}$$

com menores principais líderes $|H_1|, |H_2|, \dots, |H_n|$, como definido anteriormente. A condição suficiente de segunda ordem para um extremo é, como antes, que todos os menores principais sejam positivos (para um mínimo em z) e que seus sinais se alternem devidamente (para um máximo em z), sendo o primeiro negativo.

Em resumo, então – se nos concentrarmos no teste com determinantes – temos os critérios relacionados na Tabela 11.2, que é válida para uma função objetivo de qualquer número de variáveis escolhidas. Como casos especiais, podemos ter $n = 1$ ou $n = 2$. Quando $n = 1$, a função objetivo é $z = f(x)$, e as condições para maximização, $f_1 = 0$ e $|H_1| < 0$, se reduzem a $f'(x) = 0$ e $f''(x) < 0$, exatamente como aprendemos na Seção 9.4. De modo semelhante, quando $n = 2$, a função objetivo é $z = f(x_1, x_2)$, de modo que a condição de primeira ordem para um máximo é $f_1 = f_2 = 0$, ao passo que a condição suficiente de segunda ordem torna-se

$$f_{11} < 0 \quad \text{e} \quad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = f_{11} f_{22} - f_{12}^2 > 0$$

que é uma mera confirmação das informações apresentadas na Tabela 11.1.

Condição	Máximo	Mínimo
Condição necessária de primeira ordem	$f_1 = f_2 = \dots = f_n = 0$	$f_1 = f_2 = \dots = f_n = 0$
Condição suficiente de segunda ordem [†]	$ H_1 < 0; H_2 > 0;$ $ H_3 < 0; \dots; (-1)^n H_n > 0$	$ H_1 , H_2 , \dots, H_n > 0$

[†] Aplicável somente após a condição necessária de primeira ordem ter sido satisfeita.

TABELA 11.2
Teste com Determinantes para Extremo Relativo:
 $z = f(x_1, x_2, \dots, x_n)$

EXERCÍCIO 11.4

Encontre os valores extremos, se houver, das quatro funções seguintes. Verifique se são mínimos ou máximos pelo teste com determinantes.

1. $z = x_1^2 + 3x_2^2 - 3x_1x_2 + 4x_2x_3 + 6x_3^2$

2. $z = 29 - (x_1^2 + x_2^2 + x_3^2)$

3. $z = x_1x_3 + x_1^2 - x_2 + x_2x_3 + x_2^2 + 3x_3^2$

4. $z = e^{2x} + e^{-y} + e^{w^2} - (2x + 2e^w - y)$

Agora responda às seguintes perguntas referentes a matrizes hessianas e suas raízes características.

5. (a) Quais dos Problemas de 1 a 4 resultam em matrizes hessianas diagonais? Em cada um desses casos, os elementos da diagonal possuem um sinal uniforme?
- (b) O que você pode concluir sobre as raízes características de cada matriz hessiana diagonal encontrada? E sobre a condição de sinal definido de d^2z ?
- (c) Os resultados do teste da raiz característica estão de acordo com os do teste com determinantes?
6. (a) Encontre as raízes características da matriz hessiana para o Problema 3.
- (b) O que você pode concluir de seus resultados?
- (c) A sua resposta a (b) é consistente com o resultado do teste com determinantes para o Problema 3?

11.5 Condições de segunda ordem relativas à concavidade e convexidade

Condições de segunda ordem – sejam enunciadas em termos dos menores principais do determinante hessiano ou das raízes características da matriz hessiana – estão sempre relacionadas à questão de um ponto estacionário ser o pico de uma colina ou o fundo de um vale. Em outras palavras, referem-se ao modo como uma curva, superfície ou hipersuperfície (conforme o caso) se curva ao redor de um ponto estacionário. No caso de uma única variável, com $z = f(x)$, a configuração de colina (vale) é manifestada por uma curva inversa (em forma de U). Para a função de duas variáveis, $z = f(x, y)$, a configuração de colina (vale) toma a forma de uma superfície em formato de um domo (ou de uma tigela), como ilustrado na Figura 11.2a (Figura 11.2b). Quando estão presentes três ou mais variáveis de escolha, as colinas e os vales não podem mais ser representados graficamente, porém, ainda podemos imaginar “colinas” e “vales” em hipersuperfícies.

Uma função que dá origem a uma colina (vale) em todo o seu domínio é denominada uma função *côncava* (*convexa*).[†] Na presente discussão, admitiremos que o domínio é todo o R^n , onde n é o número de variáveis de escolha. Considerando que a caracterização como colina ou vale refere-se a todo o domínio, concavidade e convexidade são, é claro, conceitos globais. Se quisermos uma classificação mais minuciosa, podemos também distinguir entre concavidade e convexidade por um lado, e concavidade *estrita* e convexidade *estrita* por outro. No caso de *não-estrita*, a colina ou vale pode conter uma ou mais porções planas (em oposição a curvas), tais como segmentos de reta (em uma curva) ou segmentos de plano (em uma superfície). A presença da palavra *estrita* exclui, entretanto, tais segmentos de reta ou de plano. As duas superfícies mostradas na Figura 11.2 ilustram funções estritamente côncavas e estritamente convexas, respectivamente. A curva na Figura 6.5, por sua vez, é convexa (mostra um vale), mas não estritamente convexa (contém segmentos de reta). Uma função estritamente côncava (estritamente convexa) deve ser côncava (convexa), mas a recíproca não é verdadeira.

Em vista da associação de concavidade e concavidade estrita com uma configuração global de colina, um extremo de uma função côncava deve ser um pico – um máximo (em oposição a

[†] Se a colina (vale) aparece somente sobre um subconjunto S do domínio, diz-se que a função é *côncava* (*convexa*) em S .

um mínimo). Além do mais, esse máximo deve ser um máximo absoluto (em oposição a um máximo relativo), já que a colina abrange o domínio inteiro. Contudo, aquele máximo absoluto pode não ser único, porque podem ocorrer vários máximos se a colina tiver um topo horizontal plano. Esta última possibilidade só pode ser descartada quando especificamos concavidade estrita. Pois é apenas nesse caso que um pico consistirá em um único ponto e o máximo absoluto será *único*. Um máximo absoluto único (não-único) também é denominado um máximo absoluto *forte* (*fraco*).

Por raciocínio análogo, um extremo de uma função *convexa* deve ser um mínimo absoluto (ou global), que pode não ser único. Mas um extremo de uma função *estritamente convexa* deve ser um mínimo absoluto único.

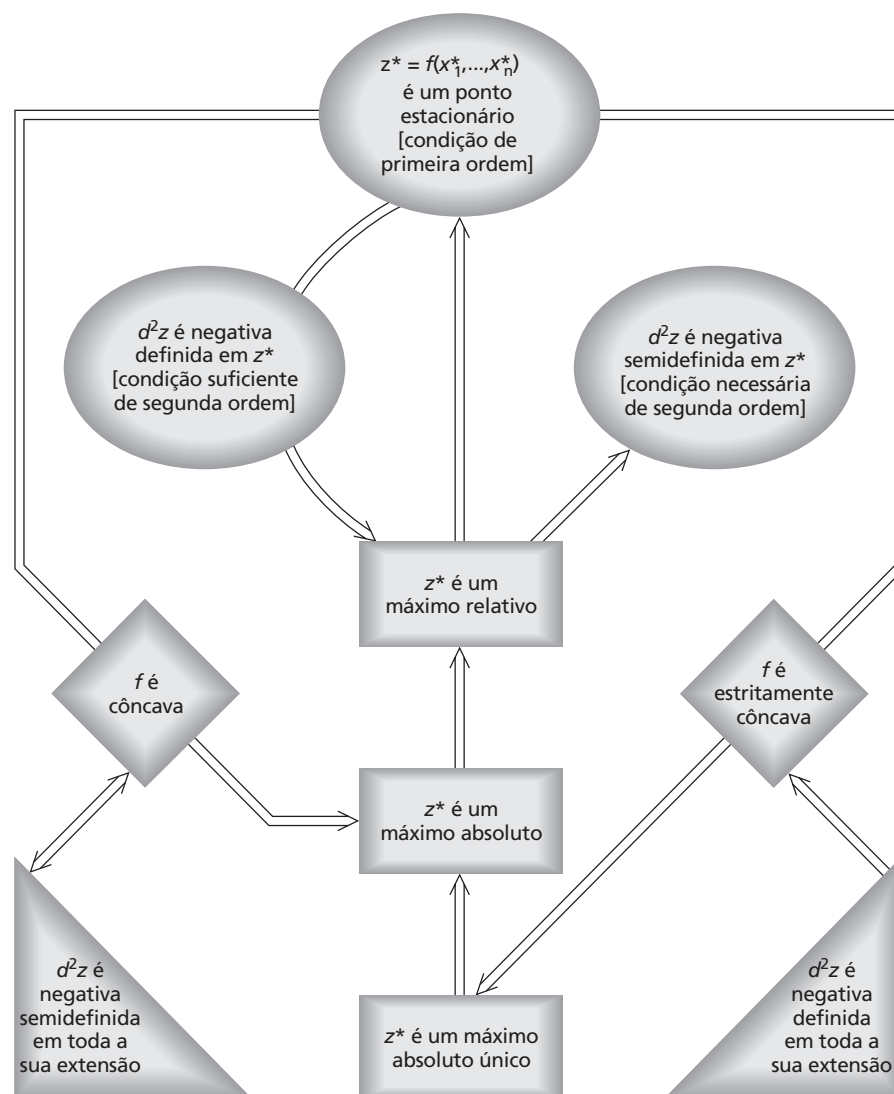
Nos parágrafos precedentes, admitimos que as propriedades de concavidade e convexidade são de escopo global. Se forem válidas somente para uma porção da curva ou da superfície (somente para um subconjunto S do domínio), então o máximo e o mínimo associados são relativos (ou locais) para aquele subconjunto do domínio, uma vez que não podemos ter certeza da situação fora do subconjunto S . Quando discutimos a condição de sinal definido de d^2z (ou da matriz hessiana H), avaliamos os menores principais líderes do determinante hessiano somente no ponto estacionário. Limitando assim a verificação de configuração de colina ou vale a uma pequena vizinhança do ponto estacionário, poderíamos discutir apenas máximos e mínimos *relativos*. Mas pode acontecer de d^2z ter um sinal definido em toda sua extensão, independentemente de onde os menores principais líderes são avaliados. Nesse caso, a colina ou vale cobriria todo o domínio e o máximo ou mínimo encontrado seria de natureza absoluta. Mais especificamente, se d^2z for negativa (positiva) semidefinida *em toda a sua extensão*, a função $z = f(x_1, x_2, \dots, x_n)$ deve ser côncava (convexa) e, se d^2z for negativa (positiva) definida *em toda a sua extensão*, a função f deve ser estritamente côncava (estritamente convexa).

A discussão precedente está resumida na Figura 11.5 para uma função $z = f(x_1, x_2, \dots, x^n)$ continuamente diferenciável duas vezes. Por questão de clareza, vamos nos concentrar exclusivamente em concavidade e máximo; contudo, as relações apresentadas permanecerão válidas se as palavras *côncava*, *negativa* e *máximo* forem substituídas por *convexa*, *positiva* e *mínimo*. Para ler a Figura 11.5, lembre-se que o símbolo $\Rightarrow$ (aqui alongado e até mesmo curvo), significa “implica”. Quando esse símbolo se estende de uma figura fechada (digamos, um retângulo) para outra (digamos, uma oval), significa que a primeira implica (é suficiente para) a última; significa também que a última é necessária para a primeira. E quando o símbolo $\Rightarrow$ se estender de uma figura fechada para uma segunda e para uma terceira, significa que a primeira, quando acompanhada da segunda, implica a terceira.

Segundo essa perspectiva, a coluna do meio na Figura 11.5, lida de cima para baixo, afirma que a condição de primeira ordem é necessária para z^* ser um máximo relativo, e que o status de máximo relativo de z^* , por sua vez, é necessário para que z^* seja um máximo absoluto e assim por diante. Alternativamente, lendo a coluna de baixo para cima, vemos que o fato de z^* ser um máximo absoluto único é suficiente para assinalar z^* como um máximo absoluto, e o status de máximo absoluto de z^* , por sua vez, é suficiente para que z^* seja um máximo relativo e assim por diante. As três ovals do topo têm a ver com as condições de primeira e segunda ordem no ponto estacionário z^* . Por conseguinte, elas se referem apenas a um máximo relativo. Os losangos e triângulos na parte inferior, por sua vez, descrevem propriedades globais que nos habilitam a tirar conclusões sobre um máximo absoluto. Note que, conquanto nossa discussão anterior indicasse somente que a condição de negativa semidefinida em toda a extensão de d^2z é *suficiente* para a concavidade da função f , adicionamos na Figura 11.5 a informação de que a condição também é *necessária*. Por outro lado, a propriedade mais forte da condição de negativa definida em toda a extensão de d^2z é *suficiente*, mas não *necessária*, para a concavidade estrita de f — porque a concavidade estrita de f é compatível com uma d^2z de valor zero em um ponto estacionário.

A mensagem mais importante transmitida pela Figura 11.5, contudo, está nos dois símbolos $\Rightarrow$ estendidos que passam pelos dois losangos. O da esquerda afirma que, dada uma função objetivo *côncava*, qualquer ponto estacionário pode ser imediatamente identificado como um máximo absoluto. Prosseguindo, vemos que o losango da direita indica que, se a função objetivo for *estritamente côncava*, o ponto estacionário deve ser, de fato, um máximo absoluto único. Em qualquer dos casos, uma vez satisfeita a condição de primeira ordem, a concavidade ou a concavidade

FIGURA 11.5



estrita substitui efetivamente a condição de segunda ordem como condição suficiente para um máximo – melhor dizendo, para um máximo absoluto. O poder dessa nova condição suficiente fica claro quando lembramos que pode acontecer de d^2z ser igual a zero em um pico, o que faz com que a condição suficiente de segunda ordem falhe. A concavidade, ou a concavidade estrita, entretanto, pode dar conta até mesmo desses picos difíceis, porque garante que uma condição suficiente de ordem mais alta seja satisfeita, mesmo que a de segunda ordem não seja. É por essa razão que muitas vezes os economistas supõem a concavidade desde o início, quando a intenção é formular um modelo de maximização com uma função objetivo *geral* (e, de forma semelhante, muitas vezes supõem a convexidade no caso de um modelo de minimização). Pois, então, basta aplicar a condição de primeira ordem. Todavia, observe que, se for usada uma função objetivo *específica*, a propriedade de concavidade ou convexidade já não pode mais ser simplesmente admitida. Ao contrário, ela deve ser verificada.

Verificação de concavidade e convexidade

Concavidade e convexidade, estrita ou não-estrita, podem ser definidas (e verificadas) de diversas maneiras. Em primeiro lugar, vamos apresentar uma definição geométrica de concavidade e convexidade para uma função de duas variáveis $z = f(x_1, x_2)$, semelhante à versão de uma variável discutida na Seção 9.3:

A função $z = f(x_1, x_2)$ é *côncava* (*convexa*) se, e somente se, para qualquer par de pontos distintos M e N em seu gráfico – uma superfície – um segmento de reta MN estiver *sobre* ou *abaixo* (*acima*) da superfície. A função é *estritamente côncava* (*estritamente convexa*) se, e somente se, o segmento de reta MN estiver inteiramente *abaixo* (*acima*) da superfície, exceto em M e N .

O caso de uma função estritamente côncava está ilustrado na Figura 11.6, onde M e N , dois pontos arbitrários sobre a superfície, são unidos por um segmento de reta tracejado, bem como por um arco em linha cheia, sendo que este último consiste em pontos sobre a superfície que estão diretamente acima do segmento de reta. Uma vez que a concavidade estrita requer que o segmento de reta MN esteja inteiramente abaixo do arco MN (exceto em M e N) para *qualquer* par de pontos M e N , as superfícies devem ter, tipicamente, um formato de domo. Analogamente, a superfície de uma função estritamente convexa deve ter, tipicamente, a forma de uma tigela. Quanto às funções (não estritamente) côncavas e convexas, visto que é permitido que o segmento de reta MN fique sobre a própria superfície, alguma porção da superfície ou até mesmo toda ela, pode ser um plano – achatado, em vez de curvo.

Para facilitar a generalização para o caso de n dimensões, que não pode ser representado graficamente, a definição geométrica precisa ser traduzida em uma versão algébrica equivalente. Voltando à Figura 11.6, sejam $u = (u_1, u_2)$ e $v = (v_1, v_2)$ quaisquer dois pares ordenados distintos (vetores de dimensão 2) no domínio de $z = f(x_1, x_2)$. Então, os valores de z (altura da superfície) correspondentes a eles serão $f(u) = f(u_1, u_2)$ e $f(v) = f(v_1, v_2)$, respectivamente. Supusemos que as variáveis podem assumir todos os valores reais, de modo que, se u e v estiverem no domínio, então todos os pontos sobre o segmento de reta uv também estão no domínio. Agora, cada ponto sobre o citado segmento de reta tem as características de uma “média ponderada” de u e v . Assim, podemos denotar esse segmento de reta $\theta u + (1 - \theta)v$, onde θ (a letra grega teta) – diferentemente de u e v – é um escalar (variável) cuja faixa de valores é $0 \leq \theta \leq 1$.[†] Pelo mesmo raciocínio, o segmento de reta MN , representando o conjunto de todas as médias ponderadas de $f(u)$ e $f(v)$, pode ser expresso por $\theta f(u) + (1 - \theta)f(v)$, com θ novamente variando de 0 a 1. E o arco MN ao longo da superfície? Uma vez que esse arco mostra os valores da função f avaliados nos vários pontos sobre o segmento de reta uv , ela pode ser escrita simplesmente como $f[\theta u + (1 - \theta)v]$. Usando essas expressões, agora podemos enunciar a seguinte definição algébrica:

Uma função f é $\begin{cases} \text{côncava} \\ \text{convexa} \end{cases}$ se, e somente se, para qualquer par de pontos distintos u e v no domínio de f , e para $0 < \theta < 1$,

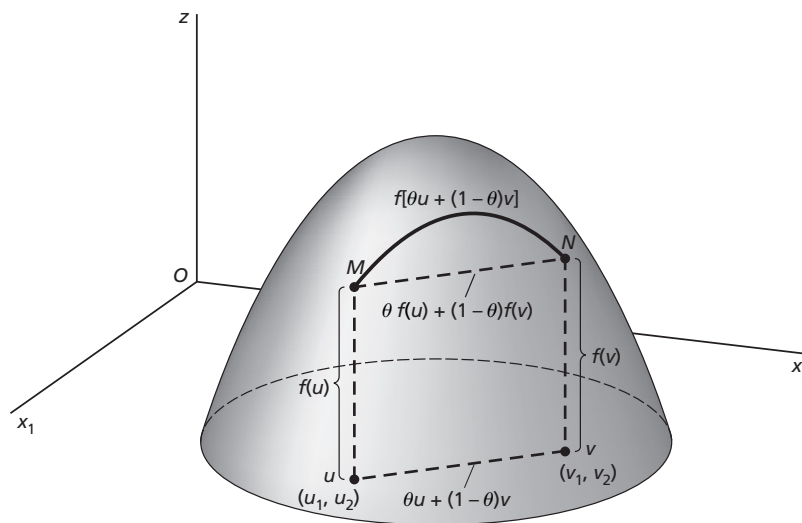


FIGURA 11.6

[†] A expressão da média ponderada $\theta u + (1 - \theta)v$ para qualquer valor específico de θ entre 0 e 1 é tecnicamente conhecido como uma *combinação convexa* dos dois vetores, u e v . Vamos deixar uma explanação mais detalhada disso para mais adiante nesta seção, mas podemos observar aqui que, quando $\theta = 0$, a expressão dada se reduz ao vetor v e que, de modo semelhante, quando $\theta = 1$, a expressão se reduz ao vetor u . Um valor intermediário de θ , por outro lado, nos dá uma média dos dois vetores, u e v .

$$\underbrace{\theta f(u) + (1 - \theta)f(v)}_{\text{altura do segmento de reta}} \begin{cases} \leq \\ \geq \end{cases} \underbrace{f[\theta u + (1 - \theta)v]}_{\text{altura do arco}} \quad (11.20)$$

Note que, para excluir da comparação de alturas os dois pontos das extremidades M e N , restringimos θ apenas ao intervalo aberto $(0, 1)$.

Essa definição pode ser adaptada com facilidade para concavidade e convexidade *estritas* mudando se as desigualdades fracas $\leq$ e $\geq$ para as desigualdade estritas $<$ e $>$, respectivamente. A vantagem da definição algébrica é que ela pode ser aplicada a uma função de qualquer número de variáveis, pois os vetores u e v na definição podem perfeitamente ser interpretados como vetores de dimensão n , em vez de vetores de dimensão 2.

De (11.20), podemos deduzir, com razoável simplicidade, os três teoremas seguintes sobre concavidade e convexidade. Eles serão enunciados em termos de funções $f(x)$ e $g(x)$, mas x pode ser interpretado como um vetor de variáveis; isto é, os teoremas são válidos para qualquer número de variáveis.

Teorema I (função linear) Se $f(x)$ for uma função linear, então ela é uma função côncava, bem como uma função convexa, mas não estritamente.

Teorema II (negativa de uma função) Se $f(x)$ for uma função côncava, então $-f(x)$ é uma função convexa, e vice-versa. De modo semelhante, se $f(x)$ for uma função estritamente côncava, então $-f(x)$ é uma função estritamente convexa, e vice-versa.

Teorema III (soma de funções) Se $f(x)$ e $g(x)$ forem ambas funções côncavas (convexas), então $f(x) + g(x)$ também é uma função côncava (convexa). Se $f(x)$ e $g(x)$ forem ambas côncavas (convexas) e, além disso, qualquer delas, ou ambas, for estritamente côncava (estritamente convexa), então $f(x) + g(x)$ é estritamente côncava (convexa).

O Teorema I surge do fato de que a representação gráfica de uma função linear é uma linha reta, um plano ou um hiperplano, de modo que o “segmento de reta MN ” sempre coincide com o “arco MN ”. Por consequência, a parte da igualdade das duas desigualdades fracas em (11.20) é simultaneamente satisfeita, o que faz com que a função se qualifique como côncava e também como convexa. Contudo, visto que a função não cumpre a parte de desigualdade estrita da definição, a função linear não é nem estritamente côncava nem estritamente convexa.

Subjacente ao Teorema II está o fato de as que definições de concavidade e convexidade são diferentes apenas no sentido de desigualdade. Suponha que $f(x)$ é côncava; então

$$\theta f(u) + (1 - \theta) f(v) \leq f[\theta u + (1 - \theta) v]$$

Multiplicando tudo por -1 , e invertendo devidamente o sentido da desigualdade, obtemos

$$\theta[-f(u)] + (1 - \theta)[-f(v)] \geq -f[\theta u + (1 - \theta) v]$$

Contudo, isso é, precisamente, a condição para $-f(x)$ ser convexa. Assim, o teorema é provado para o caso de $f(x)$ côncava. A interpretação geométrica desse resultado é muito simples: a imagem especular de uma colina com referência ao plano ou hiperplano base é um vale. O caso oposto pode ser provado de modo semelhante.

Para entender a razão por trás do Teorema III, suponha que $f(x)$ e $g(x)$ são ambas côncavas. Então, valem as duas desigualdades seguintes:

$$\theta f(u) + (1 - \theta) f(v) \leq f[\theta u + (1 - \theta) v] \quad (11.21)$$

$$\theta g(u) + (1 - \theta) g(v) \leq g[\theta u + (1 - \theta) v] \quad (11.22)$$

Somando ambas, obtemos uma nova desigualdade

$$\begin{aligned} &\theta[f(u) + g(u)] + (1 - \theta)[f(v) + g(v)] \\ &\leq f[\theta u + (1 - \theta) v] + g[\theta u + (1 - \theta) v] \end{aligned} \quad (11.23)$$

Mas isso é, exatamente, a condição para $[f(x) + g(x)]$ ser côncava. Assim, o teorema é provado para o caso da concavidade. A prova para convexidade é semelhante.

Passando para a segunda parte do Teorema III, seja $f(x)$ *estritamente* côncava. Então, (11.21) torna-se uma desigualdade *estrita*:

$$\theta f(u) + (1 - \theta) f(v) < f[\theta u + (1 - \theta)v] \quad (11.21')$$

Adicionando essa expressão a (11.22), constatamos que a soma das expressões do lado esquerdo dessas duas desigualdades são *estritamente* menores que a soma das expressões do lado direito, independentemente de o sinal $<$ ou o de o sinal $=$ valer em (11.22). Isso significa que (11.23) agora também se torna uma desigualdade *estrita* e, por conseguinte, forma $[f(x) + g(x)]$ *estritamente* côncava. Além disso, a mesma conclusão surge *a fortiori*, se fizermos $g(x)$ estritamente côncava juntamente com $f(x)$, isto é, se (11.22) for convertida em uma desigualdade estrita juntamente com (11.21). Isso prova a segunda parte do teorema para o caso da côncava. A prova para o caso da convexa é semelhante.

Esse teorema, que também é válido para uma soma de mais de duas funções côncavas (convexas), às vezes pode se revelar útil porque ele possibilita a compartimentalização da tarefa de verificar a concavidade ou a convexidade de uma função que consiste em termos aditivos. Se constatarmos que os termos aditivos são individualmente côncavos (convexos), isso seria suficiente para que a função soma seja côncava (convexa).

EXEMPLO 1 Verifique a concavidade ou convexidade de $z = x_1^2 + x_2^2$. Para aplicar (11.20), sejam $u = (u_1, u_2)$ e $v = (v_1, v_2)$ dois pontos distintos quaisquer no domínio. Então, temos

$$\begin{aligned} f(u) &= f(u_1, u_2) = u_1^2 + u_2^2 \\ f(v) &= f(v_1, v_2) = v_1^2 + v_2^2 \\ \text{e} \quad f[\theta u + (1 - \theta)v] &= f \left[\underbrace{\theta u_1 + (1 - \theta)v_1}_{\text{valor de } x_1}, \underbrace{\theta u_2 + (1 - \theta)v_2}_{\text{valor de } x_2} \right] \\ &= [\theta u_1 + (1 - \theta)v_1]^2 + [\theta u_2 + (1 - \theta)v_2]^2 \end{aligned}$$

Substituindo essas expressões em (11.20), subtraindo a expressão do lado direito da expressão do lado esquerdo e reunindo termos, constatamos que a diferença entre elas é

$$\begin{aligned} &\theta(1 - \theta)(u_1^2 + u_2^2) + \theta(1 - \theta)(v_1^2 + v_2^2) - 2\theta(1 - \theta)(u_1v_1 + u_2v_2) \\ &= \theta(1 - \theta)[(u_1 - v_1)^2 + (u_2 - v_2)^2] \end{aligned}$$

Uma vez que θ é uma fração positiva, $\theta(1 - \theta)$ deve ser positiva. Além disso, visto que (u_1, u_2) e (v_1, v_2) são pontos distintos, de modo que $u_1 \neq v_1$ ou $u_2 \neq v_2$ (ou ambos), a expressão entre colchetes também deve ser positiva. Assim, a desigualdade estrita $>$ vale em (11.20), e $z = x_1^2 + x_2^2$ é estritamente convexa.

Alternativamente, podemos verificar os termos x_1^2 e x_2^2 em separado. Visto que cada um deles é individualmente estritamente convexo, sua soma também é estritamente convexa.

Como essa função é estritamente convexa, ela possui um único mínimo absoluto. É fácil verificar que esse mínimo é $z^* = 0$, atingido em $x_1^* = x_2^* = 0$, e que ele é, de fato, absoluto e único porque qualquer par ordenado $(x_1, x_2) \neq (0, 0)$ resulta em um valor de z maior que zero.

EXEMPLO 2 Verifique a concavidade ou convexidade de $z = -x_1^2 - x_2^2$. Essa função é a negativa da função no Exemplo 1. Assim, pelo Teorema II, ela é estritamente côncava.

EXEMPLO 3 Verifique a concavidade ou convexidade de $z = (x + y)^2$. Embora as variáveis sejam denotadas por x e y em vez de x_1 e x_2 , ainda assim podemos fazer com que $u = (u_1, u_2)$ e $v = (v_1, v_2)$ denotem dois pontos distintos no domínio, sendo que o índice i refere-se à i -ésima variável. Então, temos

$$f(u) = f(u_1, u_2) = (u_1 + u_2)^2$$

$$f(v) = f(v_1, v_2) = (v_1 + v_2)^2$$

$$\begin{aligned} e \quad f[\theta u + (1 - \theta)v] &= [\theta u_1 + (1 - \theta)v_1 + \theta u_2 + (1 - \theta)v_2]^2 \\ &= [\theta(u_1 + u_2) + (1 - \theta)(v_1 + v_2)]^2 \end{aligned}$$

Substituindo essas expressões em (11.20), subtraindo a expressão do lado direito da expressão do lado esquerdo e simplificando, constatamos que a diferença entre elas é

$$\begin{aligned} \theta(1 - \theta)(u_1 + u_2)^2 - 2\theta(1 - \theta)(u_1 + u_2)(v_1 + v_2) + \theta(1 - \theta)(v_1 + v_2)^2 \\ = \theta(1 - \theta)[(u_1 + u_2) - (v_1 + v_2)]^2 \end{aligned}$$

Como no Exemplo 1, $\theta(1 - \theta)$ é positiva. O quadrado da expressão entre colchetes é não negativo (desta vez, não podemos excluir o zero). Assim, a desigualdade $\geq$ vale em (11.20), e a função $(x + y)^2$ é convexa, embora não estritamente.

De acordo com isso, essa função tem um mínimo absoluto que pode não ser único. É fácil verificar que o mínimo absoluto é $z^* = 0$, atingido sempre que $x^* + y^* = 0$. Fica claro que esse é um mínimo absoluto a partir do fato de que, sempre que $x + y \neq 0$, z será maior de que $z^* = 0$. Esse mínimo não é único devido ao fato de que um número infinito de pares (x^*, y^*) pode satisfazer a condição $x^* + y^* = 0$.

Funções diferenciáveis

Como enunciado em (11.20), a definição de concavidade e convexidade não usa nenhuma derivada e, assim, não requer diferenciabilidade. Se a função *for* diferenciável, contudo, concavidade e convexidade também podem ser definidas em termos de suas derivadas primeiras. No caso de uma só variável, a definição é:

Uma função diferenciável $f(x)$ é $\begin{cases} \text{côncava} \\ \text{convexa} \end{cases}$ se, e somente se, para qualquer ponto dado u e qualquer outro ponto v no domínio,

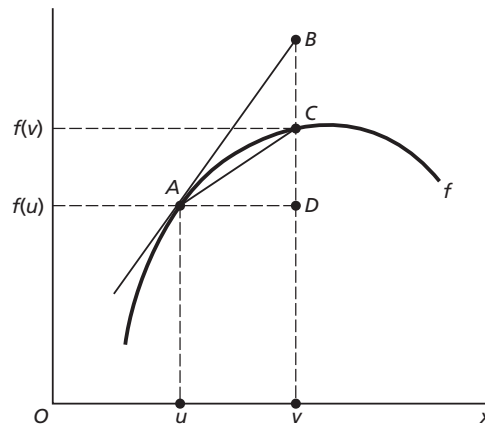
$$f(v) \begin{cases} \leq \\ \geq \end{cases} f(u) + f'(u)(v - u) \quad (11.24)$$

Concavidade e convexidade serão *estritas* se as desigualdades fracas em (11.24) forem substituídas pelas desigualdades *estritas* $<$ e $>$, respectivamente. Interpretada geometricamente, essa definição retrata uma curva côncava (convexa) como uma curva que está sobre ou abaixo (acima) de todas as suas retas tangentes. Para se qualificar como uma curva estritamente côncava (estritamente convexa), por outro lado, a curva deve estar estritamente abaixo (acima) de todas as retas tangentes, exceto nos pontos de tangência.

Na Figura 11.7, seja o ponto A qualquer ponto dado sobre a curva, com altura $f(u)$ e com a reta tangente AB . Deixemos que x aumente a partir do valor u . Então, para poder formar uma colina, uma curva estritamente côncava (como a desenhada) deve se curvar afastando-se progressivamente da reta tangente AB , de modo que o ponto C , com altura $f(v)$, tem de estar abaixo do ponto B . Nesse caso, a inclinação do segmento de reta AC é menor que o da tangente AB . Se a curva *não for* estritamente côncava, por outro lado, ela pode conter um segmento de reta de modo que, por exemplo, AC pode se transformar em um segmento de reta e coincidir com o segmento de reta AB , como uma porção linear da curva. No último caso, a inclinação de AC é igual à de AB . Juntas, essas duas situações implicam que

$$\left(\text{Inclinação do segmento de reta } AC = \frac{DC}{AD} \right) = \frac{f(v) - f(u)}{v - u} \leq (\text{inclinação de } AB =) f'(u)$$

FIGURA 11.7



Quando todos os termos são multiplicados pela quantidade positiva $(v - u)$, essa desigualdade dá o resultado em (11.24) para a função côncava. O mesmo resultado pode ser obtido se considerarmos, por sua vez, valores de x menores que u .

Quando há duas ou mais variáveis independentes, a definição precisa de uma ligeira modificação:

Uma função diferenciável $f(x) = f(x_1, \dots, x_n)$ é $\begin{cases} \text{côncava} \\ \text{convexa} \end{cases}$ se, e somente se, para qualquer ponto dado $u = (u_1, \dots, u_n)$ e qualquer outro ponto $v = (v_1, \dots, v_n)$ no domínio,

$$f(v) \begin{cases} \leq \\ \geq \end{cases} f(u) + \sum_{j=1}^n f_j(u)(v_j - u_j) \quad (11.24')$$

onde $f_j(u) \equiv \partial f / \partial x_j$ está calculado em $u = (u_1, \dots, u_n)$.

Essa definição requer que o gráfico de uma função côncava (convexa) $f(x)$ esteja sobre ou abaixo (acima) de todos os seus planos tangentes ou hiperplanos. Para concavidade e convexidade *estritas*, as desigualdades fracas em (11.24') devem ser mudadas para desigualdades *estritas*, que exigiriam que o gráfico de uma função estritamente côncava (estritamente convexa) estivesse estritamente abaixo (acima) de todos os seus planos tangentes ou hiperplanos, exceto nos pontos de tangência.

Por fim, considere a função $z = f(x_1, \dots, x_n)$ que é continuamente diferenciável duas vezes. Para tal função, existem derivadas parciais de segunda ordem e, assim, d^2z está definida. A concavidade e convexidade então podem ser verificadas pelo sinal de d^2z :

Uma função continuamente diferenciável duas vezes $z = f(x_1, \dots, x_n)$ é $\begin{cases} \text{côncava} \\ \text{convexa} \end{cases}$ se, e somente se, d^2z for $\begin{cases} \text{negativa} \\ \text{positiva} \end{cases}$ semidefinida em toda a sua extensão. A função citada é estritamente $\begin{cases} \text{côncava} \\ \text{convexa} \end{cases}$ se (mas não somente se) d^2z for $\begin{cases} \text{negativa} \\ \text{positiva} \end{cases}$ definida em toda a sua extensão. (11.25)

Não esqueça que os aspectos de côncava e estritamente côncava de (11.25) já foram incorporados à Figura 11.5.

EXEMPLO 4

Verifique a concavidade ou convexidade de $z = -x^4$ pelas condições de derivada. Em primeiro lugar, aplique (11.24). No caso presente, as expressões do lado direito e do lado esquerdo daquela desigualdade são $-v^4$ e $-u^4 - 4u^3(v - u)$, respectivamente. Subtraindo a última da primeira, constatamos que a diferença é

$$\begin{aligned}
 -v^4 + u^4 + 4u^3(v-u) &= (v-u) \left(-\frac{v^4 - u^4}{v-u} + 4u^3 \right) \quad [\text{fatorando}] \\
 &= (v-u)[-v^3 + v^2u + vu^2 + u^3] + 4u^3 \quad [\text{por (7.2)}]
 \end{aligned}$$

Seria ótimo se a expressão entre parênteses fosse divisível por $(v-u)$ pois, então, poderíamos fatorar $(v-u)$ novamente e obter o termo ao quadrado $(v-u)^2$ para facilitar a avaliação do sinal. E é exatamente isso que acontece. Assim, a equação de diferença precedente pode ser escrita como

$$-(v-u)^2[v^2 + 2vu + 3u^2] = -(v-u)^2[(v+u)^2 + 2u^2]$$

Dado que $v \neq u$, o sinal dessa expressão deve ser negativo. Valendo a desigualdade estrita $<$ em (11.24), a função $z = -x^4$ é estritamente côncava, o que significa que ela tem um máximo absoluto único. Como pode ser verificado com facilidade, esse máximo é $z^* = 0$, atingido em $x^* = 0$.

Como essa função é continuamente diferenciável duas vezes, também podemos aplicar (11.25). Uma vez que há somente uma variável, (11.25) nos dá

$$d^2z = f''(x) dx^2 = -12x^2 dx^2 \quad [\text{por (11.2)}]$$

Sabemos que dx^2 é positiva (estão sendo consideradas apenas variações diferentes de zero em x); mas $-12x^2$ pode ser negativa ou zero. Assim, o melhor que podemos fazer é concluir que d^2z é *negativa semidefinida* em toda a sua extensão e que $z = -x^4$ é (não-estritamente) côncava. Essa conclusão de (11.25) é obviamente mais fraca que a obtida anteriormente de (11.24); a saber, $z = -x^4$ é estritamente côncava. O que nos limita à conclusão mais fraca, neste caso, é o mesmo culpado que faz com que o teste da derivada segunda falhe em algumas ocasiões – o fato de que d^2z pode assumir um valor zero em um ponto estacionário de uma função que sabemos ser estritamente côncava ou estritamente convexa. E é por isso, é claro, que a condição de negativa (positiva) definida de d^2z é apresentada em (11.25) como uma condição apenas suficiente, mas não necessária, para a concavidade estrita (convexidade estrita).

EXEMPLO 5

Verifique a concavidade ou convexidade de $z = x_1^2 + x_2^2$ pelas condições de derivada. Desta vez, temos de usar (11.24') em vez de (11.24). Sendo $u = (u_1, u_2)$ e $v = (v_1, v_2)$ dois pontos quaisquer no domínio, os dois lados de (11.24') são

$$\text{Lado esquerdo} = v_1^2 + v_2^2$$

$$\text{Lado direito} = u_1^2 + u_2^2 + 2u_1(v_1 - u_1) + 2u_2(v_2 - u_2)$$

Subtraindo a última da primeira e simplificando, podemos expressar a diferença entre elas como

$$v_1^2 - 2v_1u_1 + u_1^2 + v_2^2 - 2v_2u_2 + u_2^2 = (v_1 - u_1)^2 + (v_2 - u_2)^2$$

Posto que $(v_1, v_2) \neq (u_1, u_2)$, essa diferença é sempre positiva. Assim, a desigualdade estrita $>$ é válida em (11.24'), e $z = x_1^2 + x_2^2$ é estritamente convexa. Note que o presente resultado é uma mera confirmação do que havíamos constatado anteriormente no Exemplo 1.

Quanto à utilização de (11.25), visto que $f_1 = 2x_1$ e $f_2 = 2x_2$, temos

$$f_{11} = 2 > 0 \quad \text{e} \quad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

independentemente de onde estão calculadas as derivadas parciais de segunda ordem. Assim, d^2z é positiva definida em toda a sua extensão, o que satisfaz devidamente a condição suficiente para a convexidade estrita. No presente exemplo, portanto, (11.24') e (11.25) chegam à mesma conclusão.

Funções convexas versus conjuntos convexas

Agora que já esclarecemos o significado do adjetivo *convexo* quando aplicado a uma função, devemos explicar seu significado quando utilizado para descrever um *conjunto*. Embora conjuntos

convexos e funções convexas não deixem de estar relacionados, são conceitos distintos, e é importante não confundir-los.

Para facilitar a compreensão intuitiva, vamos começar com a caracterização geométrica de um conjunto convexo. Seja S um conjunto de pontos em um espaço bidimensional ou tridimensional. Se, para quaisquer dois pontos no conjunto S , o segmento de reta que liga esses dois pontos estiver contido inteiramente em S , então diz-se que S é um *conjunto convexo*. É óbvio que uma linha reta satisfaz essa definição e constitui um conjunto convexo. Por questão de convenção, um conjunto que consiste em um único ponto também é considerado um conjunto convexo, assim como o conjunto vazio (sem nenhum ponto). Há mais exemplos na Figura 11.8. O disco – ou seja, o círculo “sólido”, um círculo mais todos os pontos dentro dele – é um conjunto convexo porque a reta que liga quaisquer dois pontos do disco está inteiramente contida nele, como exemplificado por ab (que liga dois pontos na fronteira) e cd (que liga dois pontos interiores). Note, contudo, que um círculo (oco) *não* é, em si, um conjunto convexo. De modo semelhante, um triângulo, ou um pentágono, *não* é, em si, um conjunto convexo, mas sua versão sólida é. As outras duas figuras sólidas na Figura 11.8 *não* são conjuntos convexos. A figura em forma de paleta é reentrante (recortada); assim, um segmento de reta como gh não está contido inteiramente no conjunto. Além do mais, na figura em forma de chave constatamos não somente a característica de reentrância, mas também a presença de um buraco, que é mais uma causa de não-convexidade. Em termos gerais, para se qualificar como um conjunto convexo, um conjunto de pontos não pode conter nenhum buraco e sua fronteira não deve ter nenhuma reentrância (ou recorte).

A definição geométrica de convexidade também se aplica de imediato a conjuntos de pontos em um espaço tridimensional. Por exemplo, um cubo sólido é um conjunto convexo, ao passo que um cilindro oco não é. Quando está envolvido um espaço de quatro ou mais dimensões, entretanto, a interpretação geométrica fica menos óbvia. Então, precisamos recorrer à definição algébrica de conjuntos convexos.

Para tal finalidade, é útil introduzir o conceito de *combinação convexa* de vetores (pontos), que é um tipo especial de combinação linear. Uma combinação linear de dois vetores u e v pode ser escrita como

$$k_1 u + k_2 v$$

onde k_1 e k_2 são dois escalares. Quando ambos esses escalares estão contidos no intervalo fechado $[0, 1]$ e sua soma é igual a um, diz-se que a combinação linear é uma combinação convexa, e pode ser expressa como

$$\theta u + (1 - \theta)v \quad (0 \leq \theta \leq 1) \quad (11.26)$$

À guisa de ilustração, a combinação $\frac{1}{3} \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 4 \\ 9 \end{bmatrix}$ é uma combinação convexa. Em vista do fato de que esses dois multiplicadores escalares são frações positivas cuja soma é 1, tal combinação convexa pode ser interpretada como uma *média ponderada* dos dois vetores.[†]

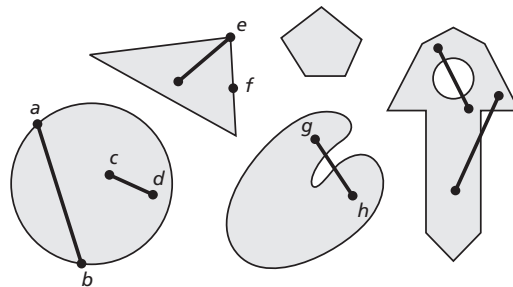


FIGURA 11.8

[†] Essa interpretação já foi utilizada anteriormente quando discutimos funções côncavas e convexas.

A característica exclusiva da combinação em (11.26) é que, para qualquer valor aceitável de θ , o vetor soma resultante está sobre o segmento de reta que liga os pontos u e v . Isso pode ser demonstrado por meio da Figura 11.9, na qual traçarmos dois vetores, $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ e $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, como dois pontos cujas coordenadas são (u_1, u_2) e (v_1, v_2) , respectivamente. Se traçarmos um outro vetor q tal que $Oquv$ forme um paralelogramo, então, temos (em virtude da discussão na Figura 4.3)

$$u = q + v \quad \text{ou} \quad q = u - v$$

Deduz-se que uma combinação convexa de vetores u e v (vamos denominá-la w) pode ser expressa em termos do vetor q , porque

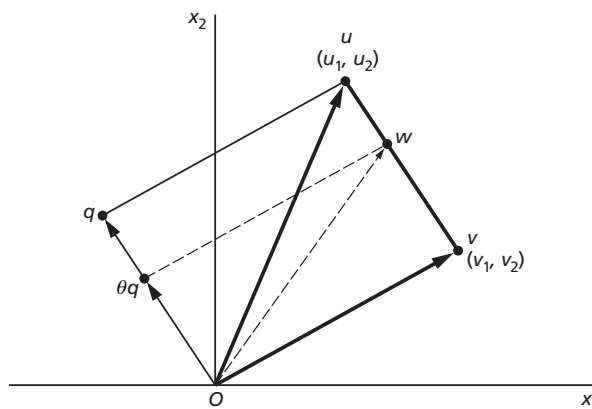
$$w = \theta u + (1 - \theta)v = \theta u + v - \theta v = \theta(u - v) + v = \theta q + v$$

Por conseguinte, para traçar o vetor w , podemos simplesmente somar θq e v pelo método familiar do paralelogramo. Se o escalar θ for uma fração positiva, o vetor θq será uma mera versão abreviada do vetor q ; assim, o vetor θq deve estar sobre o segmento de reta Oq . Portanto, somando θq e v , devemos constatar que o vetor w está sobre o segmento de reta uv , pois o novo paralelogramo, menor, nada mais é que o paralelogramo original com o lado qu deslocado para baixo. A exata localização do vetor w , é claro, vai variar de acordo com o valor do escalar θ ; quando θ varia de zero a um, a localização de w se deslocará de v para u . Assim, o conjunto de todos os pontos sobre o segmento de reta uv , incluindo os próprios u e v , corresponde ao conjunto de todas as combinações convexas dos vetores u e v .

Em vista do precedente, um conjunto convexo agora pode ser redefinido da seguinte maneira: um conjunto S é convexo se, e somente se, para quaisquer dois pontos $u \in S$ e $v \in S$, e para todo escalar $\theta \in [0, 1]$, $w = \theta u + (1 - \theta)v \in S$. Como essa definição é algébrica, ela pode ser aplicada independentemente da dimensão do espaço em que os vetores u e v estão localizados. Comparando essa definição de um conjunto convexo com a de uma função convexa em (11.20), vemos que, ainda que seja usado o mesmo adjetivo *convexo* (*a*) para ambos, o significado dessa palavra muda radicalmente de um contexto para o outro. Quando descreve uma *função*, a palavra *convexa* especifica como uma curva se comporta – ela deve formar um vale. Porém, ao descrever um *conjunto*, a palavra *convexo* especifica como os pontos do conjunto são “empacotados” – eles não devem permitir que apareça nenhum buraco e a fronteira não deve ser recortada. Assim, funções convexas e conjuntos convexas são claramente entidades matemáticas distintas.

Ainda assim, funções convexas e conjuntos convexas não deixam de estar relacionados. Uma razão é que, ao definir uma função convexa, precisamos de um conjunto convexo para o domínio. Isso porque a definição (11.20) requer que, para quaisquer dois pontos u e v no domínio, todas as combinações convexas de u e v – especificamente, $\theta u + (1 - \theta)v$, $0 \leq \theta \leq 1$ – também devem estar no domínio, o que é, evidentemente, apenas mais um modo de dizer que o domínio

FIGURA 11.9



deve ser um conjunto convexo. Para satisfazer esse requisito, adotamos anteriormente a premissa bastante forte de que o domínio consiste em todo o espaço de n dimensões (onde n é o número de variáveis de escolha), que é, de fato, um conjunto convexo. Contudo, tendo o conceito de conjuntos convexas à nossa disposição, agora podemos enfraquecer substancialmente aquela premissa – basta que adotemos como premissa que o domínio é um subconjunto convexo de R^n , e não o próprio R^n .

Há ainda um outro modo pelo qual funções convexas estão relacionadas com conjuntos convexas. Se $f(x)$ for uma função convexa, então, para qualquer constante k , ela pode dar origem a um conjunto convexo

$$S^{\leq} \equiv \{x \mid f(x) \leq k\} \quad [f(x) \text{ convexa}] \quad (11.27)$$

Isso está ilustrado na Figura 11.10a para o caso de uma única variável. O conjunto $S^{\leq}$ consiste em todos os valores x associados à parte do gráfico de $f(x)$ que está sobre ou abaixo da reta horizontal tracejada. Por conseguinte, é o segmento de reta sobre o eixo horizontal marcado pelos pontos cheios, o qual é um conjunto convexo. Note que, se o valor de k for mudado, o conjunto $S^{\leq}$ se tornará um segmento de reta diferente sobre o eixo horizontal, mas ainda será um conjunto convexo.

Avançando um pouco mais, podemos observar que, mesmo uma função *côncava* está relacionada a conjuntos convexas de modo semelhante. Primeiro, a definição de uma função côncava em (11.20), assim como o caso da função convexa, implica um domínio que é um conjunto convexo. Além disso, mesmo uma função côncava – digamos, $g(x)$ – pode gerar um conjunto convexo associado, dada alguma constante k . Essa conjunto convexo é

$$S^{\geq} \equiv \{x \mid g(x) \geq k\} \quad [g(x) \text{ côncava}] \quad (11.28)$$

no qual aparece o sinal $\geq$ em vez de $\leq$. Em termos geométricos, como mostra a Figura 11.10b para o caso de uma única variável, o conjunto $S^{\geq}$ contém todos os valores de x correspondentes à parte do gráfico de $g(x)$ que está sobre ou acima da reta horizontal tracejada. Assim, mais uma vez, ele é um segmento de reta sobre o eixo horizontal – um conjunto convexo.

Embora a Figura 11.10 ilustre especificamente o caso de uma única variável, as definições de $S^{\leq}$ e $S^{\geq}$ em (11.27) e (11.28) não estão limitadas a funções de uma única variável. Elas são igualmente válidas se interpretarmos x como um vetor, isto é, se fizermos $x = (x_1, \dots, x_n)$. Nesse caso, todavia, (11.27) e (11.28) definirão, por sua vez, conjuntos convexas no espaço de n dimensões. É importante lembrar que, conquanto a função convexa implique (11.27) e uma função côncava implique (11.28), a recíproca não é verdadeira – pois (11.27) também pode ser satisfeita por uma função não-convexa e (11.28) por uma função não-côncava. Isso é discutido com mais detalhes na Seção 12.4.

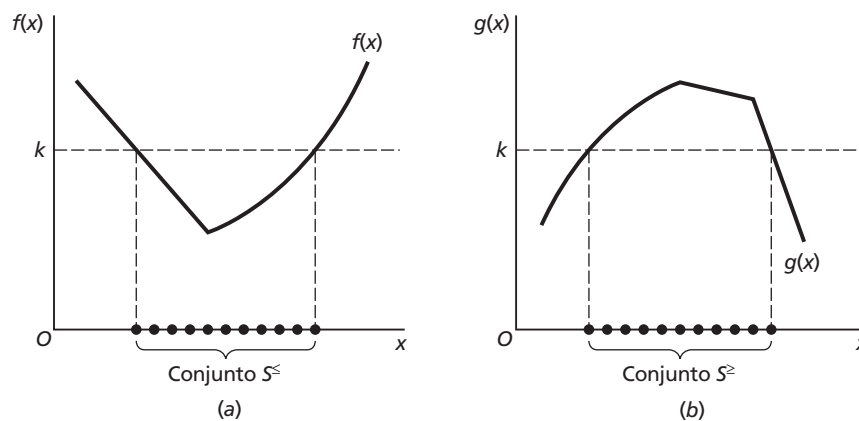


FIGURA 11.10

EXERCÍCIO 11.5

- Use (11.20) para verificar se as seguintes funções são côncavas, convexas, estritamente côncavas, estritamente convexas ou nenhuma delas:
 (a) $z = x^2$ (b) $z = x_1^2 + 2x_2^2$ (c) $z = 2x^2 - xy + y^2$
- Use (11.24) ou (11.24') para verificar se as seguintes funções são côncavas, convexas, estritamente côncavas, estritamente convexas ou nenhuma delas:
 (a) $z = -x^2$ (b) $z = (x_1 + x_2)^2$ (c) $z = -xy$
- Em vista de sua resposta ao Problema 2c, você poderia ter utilizado o Teorema III apresentado nesta seção para compartimentalizar a tarefa de verificar a função $z = 2x^2 - xy + y^2$ no Problema 1c? Justifique sua resposta.
- Os seguintes constituem conjuntos convexas no espaço tridimensional?
 (a) Uma rosquinha (b) Um pino de boliche (c) Uma bola de gude perfeita
- A equação $x^2 + y^2 = 4$ representa um círculo cujo centro é (0, 0) e cujo raio é 2.
 (a) Qual é a interpretação geométrica do conjunto $\{(x, y) \mid x^2 + y^2 \leq 4\}$?
 (b) Esse conjunto é convexo?
- Represente o gráfico de cada um dos seguintes conjuntos e indique se são convexas:
 (a) $\{(x, y) \mid y = e^x\}$ (c) $\{(x, y) \mid y \leq 13 - x^2\}$
 (b) $\{(x, y) \mid y \geq e^x\}$ (d) $\{(x, y) \mid xy \geq 1; x > 0, y > 0\}$
- Dado $u = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$ e $v = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, quais das seguintes são combinações convexas de u e v ?
 (a) $\begin{bmatrix} 7 \\ 7 \end{bmatrix}$ (b) $\begin{bmatrix} 5,2 \\ 7,6 \end{bmatrix}$ (c) $\begin{bmatrix} 6,2 \\ 8,2 \end{bmatrix}$
- Dados dois vetores u e v no espaço bidimensional, encontre e esboce:
 (a) O conjunto de todas as combinações lineares de u e v .
 (b) O conjunto de todas as combinações lineares não negativas de u e v .
 (c) O conjunto de todas as combinações convexas de u e v .
- (a) Reescreva (11.27) e (11.28) especificamente para os casos em que as funções f e g têm n variáveis independentes.
 (b) Seja $n = 2$, e seja a forma da função f uma casquinha de sorvete (na vertical), ao passo que a forma da função g é uma pirâmide. Descreva os conjuntos $S^{\leq}$ e $S^{\geq}$.

11.6 Aplicações na economia

No início deste capítulo, citamos o caso de uma empresa com vários produtos como uma ilustração do problema geral de otimização com mais de uma variável de escolha. Agora já temos conhecimento suficiente para resolver este e outros problemas de natureza semelhante.

Problema de uma empresa com vários produtos**EXEMPLO 1**

Em primeiro lugar, vamos postular uma empresa com dois produtos sob circunstâncias de concorrência pura. Uma vez que nesse tipo de concorrência os preços de ambas as mercadorias devem ser considerados exógenos, eles serão denotados por P_{10} e P_{20} , respectivamente. De acordo com isso, a função receita da empresa será

$$R_1 = P_{10}Q_1 + P_{20}Q_2$$

onde Q_i representa o nível de produção do i -ésimo produto por unidade de tempo. Vamos supor que a função custo da empresa é

$$C = 2Q_1^2 + Q_1Q_2 + 2Q_2^2$$

Note que $\partial C/\partial Q_1 = 4Q_1 + Q_2$ (o custo marginal do primeiro produto) é uma função não somente de Q_1 , mas também de Q_2 . De modo semelhante, o custo marginal do segundo produto também depende, em parte, do nível de produção do primeiro produto. Assim, de acordo com a função custo suposta, considera-se que as duas mercadorias estão tecnicamente relacionadas em termos de produção.

A função lucro dessa empresa hipotética agora pode ser escrita, de pronto, como

$$\pi = R - C = P_{10}Q_1 + P_{20}Q_2 - 2Q_1^2 - Q_1Q_2 - 2Q_2^2$$

uma função de duas variáveis de escolha (Q_1 e Q_2) e de dois parâmetros de preço. Nossa tarefa é calcular os níveis de Q_1 e Q_2 que, combinados, maximizarão π . Para essa finalidade, em primeiro lugar, vamos achar as derivadas parciais de primeira ordem da função lucro:

$$\begin{aligned}\pi_1 \left(\equiv \frac{\partial \pi}{\partial Q_1} \right) &= P_{10} - 4Q_1 - Q_2 \\ \pi_2 \left(\equiv \frac{\partial \pi}{\partial Q_2} \right) &= P_{20} - Q_1 - 4Q_2\end{aligned}\quad (11.29)$$

Igualando ambas a zero, para satisfazer a condição necessária para um máximo, obtemos duas equações simultâneas

$$\begin{aligned}4Q_1 + Q_2 &= P_{10} \\ Q_1 + 4Q_2 &= P_{20}\end{aligned}$$

que resultam na a solução única

$$Q_1^* = \frac{4P_{10} - P_{20}}{15} \quad \text{e} \quad Q_2^* = \frac{4P_{20} - P_{10}}{15}$$

Assim, se $P_{10} = 12$ e $P_{20} = 18$, por exemplo, temos $Q_1^* = 2$ e $Q_2^* = 4$, implicando um lucro ótimo $\pi^* = 48$ por unidade de tempo.

Para termos certeza de que isso realmente representa um lucro máximo, vamos verificar a condição de segunda ordem. As derivadas parciais segundas, que podem ser obtidas pela diferenciação parcial de (11.29), nos dão o seguinte hessiano:

$$|H| = \begin{vmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{vmatrix} = \begin{vmatrix} -4 & -1 \\ -1 & -4 \end{vmatrix}$$

Visto que $|H_1| = -4 < 0$ e $|H_2| = 15 > 0$, a matriz hessiana (ou d^2z) é negativa definida, e a solução realmente maximiza o lucro. De fato, uma vez que os sinais dos menores principais líderes não dependem de onde estão calculados, d^2z é, neste caso, negativa definida *em toda a sua extensão*. Assim, de acordo com (11.25), a função objetivo deve ser estritamente côncava, e o lucro máximo que acabamos de calcular é um máximo absoluto único.

EXEMPLO 2

Agora vamos transportar o problema do Exemplo 1 para o cenário de um mercado monopolista. Em virtude dessa nova premissa de estrutura de mercado, a função receita deve ser modificada para refletir o fato de que, agora, os preços dos dois produtos vão variar conforme seus níveis de produção (que supomos idênticos a seus níveis de vendas, sendo que o modelo não considera o acúmulo de nenhum nível de estoque). A maneira exata pela qual os preços irão variar com os níveis de produção deve ser encontrada, é claro, nas funções demanda para os dois produtos da empresa.

Suponha que as demandas que a empresa monopolista enfrenta são as seguintes:

$$\begin{aligned}Q_1 &= 40 - 2P_1 + P_2 \\ Q_2 &= 15 + P_1 - P_2\end{aligned}\quad (11.30)$$

Essas equações revelam que o *consumo* dessas duas mercadorias está relacionado; especificamente, são bens substitutos, porque um aumento no preço de uma delas aumentará a demanda pela outra. Dadas essas premissas, (11.30) expressa as quantidades demandadas Q_1 e Q_2 como

funções de preços, mas, para nossa presente finalidade, será mais conveniente expressar os preços P_1 e P_2 em termos dos volumes de vendas Q_1 e Q_2 , isto é, ter funções receita média para os dois produtos. Visto que (11.30) pode ser reescrita como

$$-2P_1 + P_2 = Q_1 - 40$$

$$P_1 - P_2 = Q_2 - 15$$

podemos (considerando Q_1 e Q_2 como parâmetros) aplicar a regra de Cramer para resolver para P_1 e P_2 como se segue:

$$P_1 = 55 - Q_1 - Q_2$$

$$P_2 = 70 - Q_1 - 2Q_2 \quad (11.30')$$

Essas expressões constituem as funções receita média desejadas, já que $P_1 \equiv AR_1$ e $P_2 \equiv AR_2$. Por consequência, a função receita total da empresa pode ser escrita como

$$\begin{aligned} R &= P_1Q_1 + P_2Q_2 \\ &= (55 - Q_1 - Q_2)Q_1 + (70 - Q_1 - 2Q_2)Q_2 \quad [\text{por (11.30')}] \\ &= 55Q_1 + 70Q_2 - 2Q_1Q_2 - Q_1^2 - 2Q_2^2 \end{aligned}$$

Se admitirmos novamente que a função custo total é

$$C = Q_1^2 + Q_1Q_2 + Q_2^2$$

então a função lucro será

$$\pi = R - C = 55Q_1 + 70Q_2 - 3Q_1Q_2 - 2Q_1^2 - 3Q_2^2 \quad (11.31)$$

que é uma função objetivo com duas variáveis de escolha. Uma vez encontrados os níveis de produção Q_1^* e Q_2^* que maximizam o lucro, entretanto, é bem fácil achar os preços ótimos P_1^* e P_2^* por (11.30').

A função objetivo resulta nas seguintes derivadas parciais primeira e segunda:

$$\begin{aligned} \pi_1 &= 55 - 3Q_2 - 4Q_1 & \pi_2 &= 70 - 3Q_1 - 6Q_2 \\ \pi_{11} &= -4 & \pi_{12} = \pi_{21} &= -3 & \pi_{22} &= -6 \end{aligned}$$

Para satisfazer a condição de primeira ordem para um máximo de π , devemos ter $\pi_1 = \pi_2 = 0$; isto é,

$$4Q_1 + 3Q_2 = 55$$

$$3Q_1 + 6Q_2 = 70$$

Assim, a solução para os níveis de produção (por unidade de tempo) é:

$$(Q_1^*, Q_2^*) = \left(8, 7\frac{2}{3}\right)$$

Substituindo esse resultado em (11.30') e (11.31), respectivamente, obtemos

$$P_1^* = 39\frac{1}{3} \quad P_2^* = 46\frac{2}{3} \quad \text{e} \quad \pi^* = 488\frac{1}{3} \quad (\text{por unidade de tempo})$$

Considerando que o hessiano é $\begin{vmatrix} -4 & -3 \\ -3 & -6 \end{vmatrix}$, temos

$$|H_1| = -4 < 0 \quad \text{e} \quad |H_2| = 15 > 0$$

de modo que o valor de π^* realmente representa o lucro máximo. Aqui, os sinais dos menores principais líderes são novamente independentes de onde são calculados. Assim, a matriz hessiana é negativa definida em toda a sua extensão, implicando que a função objetivo é estritamente côncava e que ela tem um único máximo absoluto.

Discriminação de preços

Mesmo em uma empresa que produz somente um produto pode surgir um problema de otimização envolvendo duas ou mais variáveis de escolha. Esse seria o caso, por exemplo, de uma empresa monopolista que vende um só produto em dois ou mais mercados separados (por exemplo, mercados interno e externo) e, portanto, precisa decidir quais quantidades (Q_1 , Q_2 etc.) devem ser fornecidas aos respectivos mercados de modo a maximizar o lucro. Em geral, as condições de demanda dos diversos mercados serão diferentes e, se as elasticidades também forem diferentes nos vários mercados, a maximização do lucro acarretará a prática da discriminação de preços. Vamos derivar matematicamente essa conclusão familiar.

EXEMPLO 3

Para mudar o cenário, desta vez vamos usar três variáveis de escolha, isto é, vamos admitir três mercados separados. Vamos também trabalhar com funções gerais em vez de numéricas. De acordo com isso, admitiremos simplesmente que nossa empresa monopolista tem as seguintes funções receita total e custo total:

$$R = R_1(Q_1) + R_2(Q_2) + R_3(Q_3)$$

$$C = C(Q) \quad \text{onde} \quad Q = Q_1 + Q_2 + Q_3$$

Note que o símbolo R_i representa aqui a função receita do i -ésimo mercado, e não uma derivada no sentido de f_i . Cada uma dessas funções receita implica, naturalmente, uma estrutura de demanda particular que, em geral, será diferente das prevalecentes nos outros dois mercados. No lado do custo, por sua vez, é postulada somente uma função custo, visto que uma única empresa está produzindo para todos os três mercados. Em vista do fato de que $Q = Q_1 + Q_2 + Q_3$, o custo total C também é, basicamente, uma função de Q_1 , Q_2 e Q_3 , as quais constituem as variáveis de escolha do modelo. É claro que podemos reescrever $C(Q)$ como $C(Q_1 + Q_2 + Q_3)$. Porém, é preciso observar que, embora esta última versão contenha três variáveis independentes, ainda assim deve-se considerar que a função tem apenas um único argumento, porque a soma de Q_i é, na verdade, uma única entidade. Ao contrário, se a função aparecer na forma $C(Q_1, Q_2, Q_3)$, então poderão ser contados tantos argumentos quantos forem as variáveis independentes.

Agora a função lucro é

$$\pi = R_1(Q_1) + R_2(Q_2) + R_3(Q_3) - C(Q)$$

com as derivadas parciais primeiras $\pi_i \equiv \partial\pi/\partial Q_i$ (para $i = 1, 2, 3$) como se segue:[†]

$$\begin{aligned} \pi_1 &= R_1'(Q_1) - C'(Q) \frac{\partial Q}{\partial Q_1} = R_1'(Q_1) - C'(Q) & \left[\text{visto que } \frac{\partial Q}{\partial Q_1} = 1 \right] \\ \pi_2 &= R_2'(Q_2) - C'(Q) \frac{\partial Q}{\partial Q_2} = R_2'(Q_2) - C'(Q) & \left[\text{visto que } \frac{\partial Q}{\partial Q_2} = 1 \right] \\ \pi_3 &= R_3'(Q_3) - C'(Q) \frac{\partial Q}{\partial Q_3} = R_3'(Q_3) - C'(Q) & \left[\text{visto que } \frac{\partial Q}{\partial Q_3} = 1 \right] \end{aligned} \quad (11.32)$$

Igualando essas expressões a zero simultaneamente, temos

$$C'(Q) = R_1'(Q_1) = R_2'(Q_2) = R_3'(Q_3)$$

[†] Note que, para achar $\partial C/\partial Q_i$, a regra da cadeia é utilizada:

$$\frac{\partial C}{\partial Q_i} = \frac{dC}{dQ} \frac{\partial Q}{\partial Q_i}$$

Isto é,

$$MC = MR_1 = MR_2 = MR_3$$

Assim, os níveis de Q_1 , Q_2 e Q_3 devem ser escolhidos de modo que a receita marginal em cada mercado seja igual ao custo marginal da produção total Q .

Para perceber as implicações dessa condição no que diz respeito à discriminação de preços, em primeiro lugar vamos descobrir como a MR em qualquer mercado está especificamente relacionada com o preço naquele mercado. Visto que a receita em cada mercado é $R_i = P_i Q_i$, deduz-se que a receita marginal deve ser

$$\begin{aligned} MR_i &\equiv \frac{dR_i}{dQ_i} = P_i \frac{dQ_i}{dQ_i} + Q_i \frac{dP_i}{dQ_i} \\ &= P_i \left(1 + \frac{dP_i}{dQ_i} \frac{Q_i}{P_i} \right) = P_i \left(1 + \frac{1}{\varepsilon_{di}} \right) \quad [\text{por (8.4)}] \end{aligned}$$

onde ε_{di} , a elasticidade pontual de demanda no i -ésimo mercado, normalmente é negativa. Por consequência, a relação entre MR_i e P_i pode ser expressa de modo alternativo pela equação

$$MR_i = P_i \left(1 - \frac{1}{|\varepsilon_{di}|} \right) \quad (11.33)$$

Lembre-se que $|\varepsilon_{di}|$ é, em geral, uma função de P_i , de modo que, quando Q_i^* é escolhida e, por isso, P_i^* é especificado, $|\varepsilon_{di}|$ também assumirá um valor específico que pode ser maior que um, menor que um ou igual a um. Mas se $|\varepsilon_{di}| < 1$ (demanda inelástica em um ponto), então sua recíproca será maior que um e a expressão entre parênteses em (11.33) será negativa, o que implica um valor negativo para MR_i . De modo semelhante, se $|\varepsilon_{di}| = 1$ (elasticidade unitária), então MR_i assumirá um valor zero. Considerando que o MC de uma empresa é positivo, a condição de primeira ordem $MC = MR_i$ requer que a empresa opere em um nível positivo de MR_i . Portanto, os níveis de vendas Q_i escolhidos pela empresa devem ser tais que a elasticidade pontual de demanda correspondente em cada mercado seja maior que um.

A condição de primeira ordem $MR_1 = MR_2 = MR_3$ agora pode ser traduzida por meio de (11.33), para o seguinte:

$$P_1 \left(1 - \frac{1}{|\varepsilon_{d1}|} \right) = P_2 \left(1 - \frac{1}{|\varepsilon_{d2}|} \right) = P_3 \left(1 - \frac{1}{|\varepsilon_{d3}|} \right)$$

Dessas expressões pode-se inferir imediatamente que, quanto *menor* o valor de $|\varepsilon_{di}|$ (no nível de produção escolhido) em um mercado particular, *mais alto* será o preço cobrado naquele mercado – daí, a discriminação de preços – se o que se quer é maximizar o lucro.

Para assegurar a maximização, vamos examinar a condição de segunda ordem. De (11.32), constatamos que as derivadas parciais segundas são

$$\pi_{11} = R_1'(Q_1) - C''(Q) \frac{\partial Q}{\partial Q_1} = R_1'(Q_1) - C''(Q)$$

$$\pi_{22} = R_2'(Q_2) - C''(Q) \frac{\partial Q}{\partial Q_2} = R_2'(Q_2) - C''(Q)$$

$$\pi_{33} = R_3'(Q_3) - C''(Q) \frac{\partial Q}{\partial Q_3} = R_3'(Q_3) - C''(Q)$$

$$\text{e} \quad \pi_{12} = \pi_{21} = \pi_{13} = \pi_{31} = \pi_{23} = \pi_{32} = -C''(Q) \quad \left[\text{visto que } \frac{\partial Q}{\partial Q_i} = 1 \right]$$

de modo que temos (após abreviar a notação da derivada segunda)

$$|H| = \begin{vmatrix} R_1'' - C'' & -C'' & -C'' \\ -C'' & R_2'' - C'' & -C'' \\ -C'' & -C'' & R_3'' - C'' \end{vmatrix}$$

Assim, a condição suficiente de segunda ordem será devidamente satisfeita, contanto que tenhamos:

1. $|H_1| = R_1'' - C'' < 0$; isto é, a inclinação de MR_1 é menor que a inclinação de MC de toda a produção [veja a situação do ponto L na Figura 9.6c]. (Uma vez que qualquer um dos três mercados pode ser considerado o “primeiro” mercado, isso também implica, com efeito, que $R_2'' - C'' < 0$ e $R_3'' - C'' < 0$.)
2. $|H_2| = (R_1'' - C'')(R_2'' - C'') - (C'')^2 > 0$; ou, $R_1''R_2'' - (R_1'' + R_2'')C'' > 0$.
3. $|H_3| = R_1''R_2''R_3'' - (R_1''R_2'' + R_1''R_3'' + R_2''R_3'')C'' < 0$.

As duas últimas partes dessa condição não são tão fáceis de interpretar em termos econômicos quanto a primeira. Note que, se tivéssemos suposto que as funções gerais $R_i(Q_i)$ são todas côncavas e que a função geral $C(Q)$ é convexa, de modo que $-C(Q)$ é côncava, então a função lucro – a soma das funções côncavas – poderia ter sido considerada como côncava, evitando, assim, a necessidade de verificar a condição de segunda ordem.

EXEMPLO 4

Para trazer o exemplo acima para um cenário mais concreto, vamos apresentar agora uma versão numérica. Suponha que nossa empresa monopolista tenha as seguintes funções renda média específicas

$$\begin{array}{lll} P_1 = 63 - 4Q_1 & \text{de modo que} & R_1 = P_1Q_1 = 63Q_1 - 4Q_1^2 \\ P_2 = 105 - 5Q_2 & & R_2 = P_2Q_2 = 105Q_2 - 5Q_2^2 \\ P_3 = 75 - 6Q_3 & & R_3 = P_3Q_3 = 75Q_3 - 6Q_3^2 \end{array}$$

e que a função custo total é

$$C = 20 + 15Q$$

Então, as funções marginais serão

$$R_1' = 63 - 8Q_1 \quad R_2' = 105 - 10Q_2 \quad R_3' = 75 - 12Q_3 \quad C' = 15$$

Quando igualamos cada receita marginal R_i' ao custo marginal C' da produção total, constatamos que as quantidades de equilíbrio são

$$Q_1^* = 6 \quad Q_2^* = 9 \quad \text{e} \quad Q_3^* = 5$$

$$\text{Assim,} \quad Q^* = \sum_{i=1}^3 Q_i^* = 20$$

Substituindo essas soluções nas equações de receita e custo, obtemos $\pi^* = 679$ como o lucro total da operação comercial em três mercados.

Como este é um modelo específico, sem dúvida temos de verificar a condição de segunda ordem (ou a concavidade da função objetivo). Visto que as derivadas segundas são

$$R_1'' = -8 \quad R_2'' = -10 \quad R_3'' = -12 \quad C'' = 0$$

todas as três partes das condições suficientes de segunda ordem no Exemplo 3 são devidamente satisfeitas.

É fácil observar, pelas funções receita média, que a empresa deveria cobrar os preços discriminatórios $P_1^* = 39$, $P_2^* = 60$ e $P_3^* = 45$ nos três mercados. Como se pode verificar imediatamente, a elasticidade-pontual de demanda é a mais baixa de todas no segundo mercado, no qual é cobrado o preço mais alto.

Decisões de insumos de uma empresa

Em vez dos níveis de produção Q , as variáveis de escolha de uma empresa também podem aparecer sob a forma de níveis de insumos.

EXEMPLO 5 Considere uma empresa competitiva cuja função lucro é a seguinte

$$\pi = R - C = P Q - wL - rK \quad (11.34)$$

onde P = preço
 Q = produção
 L = mão-de-obra
 K = capital
 w, r = preços dos insumos L e K , respectivamente

Uma vez que a empresa funciona em um mercado competitivo, as variáveis exógenas são P, w e r (escritas aqui sem o índice zero). Há três variáveis endógenas, K, L e Q . Contudo, a produção Q é, por sua vez, uma função de K e L por meio da função produção

$$Q = Q(K, L)$$

Vamos supor que ela seja uma função de Cobb-Douglas (discutida mais a fundo na Seção 12.6) na forma

$$Q = L^\alpha K^\beta$$

onde α e β são parâmetros positivos. Se supusermos, ainda mais, retornos decrescentes de escala, então $\alpha + \beta < 1$. Por questão de simplicidade, consideraremos o caso simétrico em que $\alpha = \beta < \frac{1}{2}$

$$Q = L^\alpha K^\alpha \quad (11.35)$$

Substituindo (11.35) em (11.34), temos

$$\pi(K, L) = P L^\alpha K^\alpha - wL - rK$$

A condição de primeira ordem para a maximização do lucro é

$$\frac{\partial \pi}{\partial L} = P \alpha L^{\alpha-1} K^\alpha - w = 0$$

$$\frac{\partial \pi}{\partial K} = P \alpha L^\alpha K^{\alpha-1} - r = 0 \quad (11.36)$$

Esse sistema de equações define o L e o K ótimos para a maximização do lucro. Mas, antes, vamos verificar a condição de segunda ordem para ver se realmente temos um máximo.

O hessiano para esse problema é

$$|H| = \begin{vmatrix} \pi_{LL} & \pi_{LK} \\ \pi_{KL} & \pi_{KK} \end{vmatrix} = \begin{vmatrix} P\alpha(\alpha-1)L^{\alpha-2}K^\alpha & P\alpha^2L^{\alpha-1}K^{\alpha-1} \\ P\alpha^2L^{\alpha-1}K^{\alpha-1} & P\alpha(\alpha-1)L^\alpha K^{\alpha-2} \end{vmatrix}$$

A condição suficiente para um máximo é que $|H_1| < 0$ e $|H| > 0$:

$$\begin{aligned} |H_1| &= P\alpha(\alpha-1)L^{\alpha-2}K^\alpha < 0 \\ |H| &= P^2\alpha^2(\alpha-1)^2L^{2\alpha-2}K^{2\alpha-2} - P^2\alpha^4L^{2\alpha-2}K^{2\alpha-2} \\ &= P^2\alpha^2L^{2\alpha-2}K^{2\alpha-2}(1-2\alpha) > 0 \end{aligned}$$

Portanto, para $\alpha < \frac{1}{2}$, a condição suficiente de segunda ordem é satisfeita.

Agora podemos voltar à condição de primeira ordem para resolver para K e L ótimos. Reescrevendo a primeira equação em (11.36) de modo a isolar K , obtemos

$$\begin{aligned} P\alpha L^{\alpha-1}K^\alpha &= w \\ K &= \left(\frac{w}{P\alpha} L^{1-\alpha} \right)^{\frac{1}{\alpha}} \end{aligned}$$

Substituindo essa expressão na segunda equação de (11.36), temos

$$P\alpha L^\alpha K^{\alpha-1} - r = P\alpha L^\alpha \left[\left(\frac{W}{P\alpha} L^{1-\alpha} \right)^{\frac{1}{\alpha}} \right]^{\alpha-1} - r = 0$$

ou

$$P^{\frac{1}{\alpha}} \alpha^{\frac{1}{\alpha}} W^{(\alpha-1)/\alpha} L^{(2\alpha-1)/\alpha} = r$$

Rearrmando para resolver para L , temos, então

$$L^* = (P\alpha W^{\alpha-1} r^{-\alpha})^{1/(1-2\alpha)}$$

Aproveitando a simetria do modelo, podemos escrever rapidamente o K ótimo como

$$K^* = (P\alpha r^{\alpha-1} W^{-\alpha})^{1/(1-2\alpha)}$$

L^* e K^* são as equações de demanda de insumo da empresa.

Se substituirmos L^* e K^* na função produção, constatamos que

$$\begin{aligned} Q^* &= (L^*)^\alpha (K^*)^\alpha \\ &= (P\alpha W^{\alpha-1} r^{-\alpha})^{\alpha/(1-2\alpha)} (P\alpha r^{\alpha-1} W^{-\alpha})^{\alpha/(1-2\alpha)} \\ &= \left(\frac{\alpha^2 P^2}{W r} \right)^{\alpha/(1-2\alpha)} \end{aligned} \quad (11.37)$$

Isso nos dá uma expressão para a produção ótima como uma função das variáveis exógenas P , W , e r .

EXEMPLO 6

Vamos supor as seguintes circunstâncias: (1) Dois insumos a e b são utilizados na produção de um único produto Q de uma empresa hipotética. (2) Os preços de ambos os insumos, P_a e P_b , estão fora do controle da empresa, assim como o preço de produção P ; nós os denotaremos por P_{a0} , P_{b0} e P_0 , respectivamente. (3) O processo de produção leva t_0 anos (sendo t_0 alguma constante positiva) para ser concluído; assim, a receita de vendas tem de ser devidamente descontada antes de poder ser adequadamente comparada com o custo de produção incorrido no tempo presente. Admita-se que taxa de desconto em base contínua é dada em r_0 .

Segundo a premissa 1, podemos escrever uma função produção geral $Q = Q(a, b)$, com produtos físicos marginais Q_a e Q_b . A premissa 2 nos habilita a expressar o custo total como

$$C = P_{a0}a + P_{b0}b$$

e a receita total como

$$R = P_0 Q(a, b)$$

Antes de escrever a função lucro, entretanto, temos de descontar a receita multiplicando-a pela constante $e^{-r_0 t_0}$ – a qual, para evitar índices superiores e inferiores complicados, vamos escrever como e^{-rt} . Assim, a função lucro é

$$\pi = P_0 Q(a, b) e^{-rt} - P_{a0}a - P_{b0}b$$

na qual a e b são as únicas variáveis de escolha.

Para maximizar o lucro, é necessário que as derivadas parciais primeiras

$$\begin{aligned} \pi_a \left(\equiv \frac{\partial \pi}{\partial a} \right) &= P_0 Q_a e^{-rt} - P_{a0} \\ \pi_b \left(\equiv \frac{\partial \pi}{\partial b} \right) &= P_0 Q_b e^{-rt} - P_{b0} \end{aligned} \quad (11.38)$$

sejam ambas iguais a zero. Isso significa que

$$P_0 Q_a e^{-rt} = P_{a0} \quad \text{e} \quad P_0 Q_b e^{-rt} = P_{b0} \quad (11.39)$$

Uma vez que $P_0 Q_a$ (o preço do produto multiplicado pelo produto marginal do insumo a) representa o *valor do produto marginal do insumo a* (VMP_a), a primeira equação nos diz apenas que o valor presente de VMP_a deve ser igual ao preço do insumo a dado. A segunda equação é o mesmo pré-requisito aplicado ao insumo b .

Note que, para satisfazer (11.39), ambos os produtos marginais físicos Q_a e Q_b devem ser positivos, porque P_0 , P_{a0} , P_{b0} e e^{-rt} todos têm valores positivos. Isso tem uma importante interpretação em termos de uma *isoquanta*, definida como o locus de combinações de insumos que resultam no mesmo nível de produção. Quando representadas graficamente no plano ab , as isoquantas geralmente têm o aspecto das traçadas na Figura 11.11. Considerando que cada uma delas é pertinente a um nível de produção fixo, devemos ter, ao longo de cada isoquanta,

$$dQ = Q_a da + Q_b db = 0$$

o que implica que a inclinação de uma isoquanta pode ser expressa como

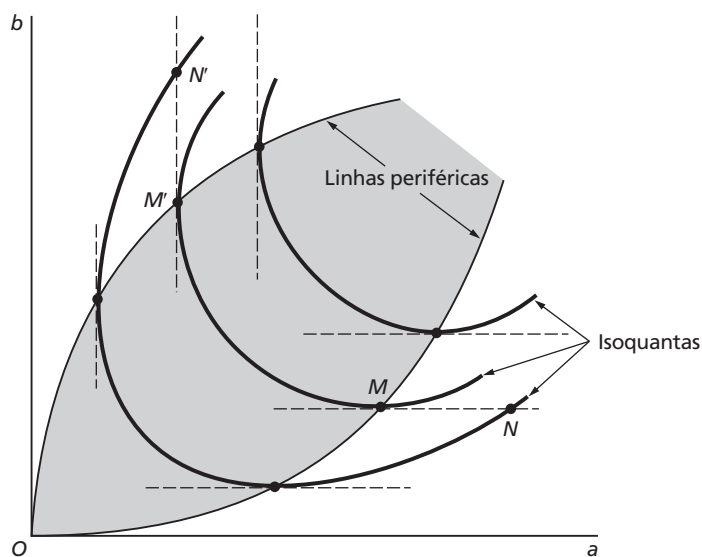
$$\frac{db}{da} = -\frac{Q_a}{Q_b} \quad \left(= -\frac{MPP_a}{MPP_b} \right) \quad (11.40)$$

Assim, para que ambas, Q_a e Q_b , sejam positivas, a empresa deverá restringir a escolha de insumos somente aos segmentos das isoquantas cujas inclinações são negativas. Por isso, na Figura 11.11, a região relevante de operação é restrita à área sombreada definida pelas denominadas linhas periféricas. Fora da área sombreada, onde as isoquantas são caracterizadas por inclinações positivas, o produto marginal de um único insumo deve ser negativo. O deslocamento da combinação de insumos em M para a combinação de insumos em N , por exemplo, indica que, se o insumo b for mantido constante, o *aumento* no insumo a nos leva a uma isoquanta *mais baixa* (uma produção menor); assim, Q_a deve ser negativa. De modo semelhante, um deslocamento de M' para N' ilustra a negatividade de Q_b . Note que, quando restringimos nossa atenção à área sombreada, cada isoquanta pode ser tomada como uma função da forma $b = \phi(a)$ porque, para cada valor admissível de a , a isoquanta determina um único valor de b .

A condição de segunda ordem gira em torno das derivadas parciais segundas de π , que podem ser obtidas de (11.38). Tendo em mente que Q_a e Q_b , na qualidade de derivadas, são, em si, funções das variáveis a e b , podemos achar π_{aa} , $\pi_{ab} = \pi_{ba}$ e π_{bb} e arranjá-las em um hessiano:

$$|H| = \begin{vmatrix} \pi_{aa} & \pi_{ab} \\ \pi_{ab} & \pi_{bb} \end{vmatrix} = \begin{vmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{vmatrix} \quad (11.41)$$

FIGURA 11.11



Para que um valor estacionário de π seja um máximo, é suficiente que

$$|H_1| < 0 \quad [\text{isto é, } \pi_{aa} < 0, \text{ que pode ocorrer se, e somente se, } Q_{aa} < 0]$$

$$|H_2| = |H| > 0 \quad [\text{isto é, } \pi_{aa}\pi_{bb} > \pi_{ab}^2, \text{ que pode ocorrer se, e somente se, } Q_{aa}Q_{bb} > Q_{ab}^2]$$

Observamos, assim, que a condição de segunda ordem pode ser testada com as derivadas π_{ij} ou com as derivadas Q_{ij} o que for mais conveniente.

O símbolo Q_{aa} denota a taxa de variação de Q_a ($\equiv \text{MPP}_a$) à medida que o insumo a varia, enquanto o insumo b permanece fixo; de modo semelhante, Q_{bb} denota a taxa de variação de Q_b ($\equiv \text{MPP}_b$) à medida que apenas o insumo b varia. Portanto, a condição suficiente de segunda ordem estipula, em parte, que os MPP de ambos os insumos sejam *decrecentes* nos níveis de insumos escolhidos a^* e b^* . Observe, entretanto, que MPP_a e MPP_b decrescentes *não* garantem a satisfação da condição de segunda ordem, porque esta última condição também envolve a grandeza de $Q_{ab} = Q_{ba}$, que mede a taxa de variação do MPP de um único insumo enquanto a quantidade do outro insumo varia.

Um exame mais profundo revela que, exatamente como a condição de primeira ordem especifica que a inclinação da isoquanta deve ser negativa na combinação de insumos escolhida (como demonstra a área sombreada da Figura 11.11), a condição suficiente de segunda ordem serve para especificar que a mesma isoquanta seja estritamente convexa na combinação de insumos escolhida. A curvatura da isoquanta está associada ao sinal da derivada segunda d^2b/da^2 . Para obter esta última, (11.40) deve ser diferenciada totalmente em relação a a , tendo em mente que ambas, Q_a e Q_b , são funções derivadas de a e b e, ainda assim, sobre uma isoquanta, b é, em si, uma função de a ; isto é,

$$Q_a = Q_a(a, b) \quad Q_b = Q_b(a, b) \quad \text{e} \quad b = \phi(a)$$

Assim, a diferenciação total prossegue do seguinte modo:

$$\frac{d^2b}{da^2} = \frac{d}{da} \left(-\frac{Q_a}{Q_b} \right) = -\frac{1}{Q_b^2} \left[Q_b \frac{dQ_a}{da} - Q_a \frac{dQ_b}{da} \right] \quad (11.42)$$

Visto que b é uma função de a sobre a isoquanta, a fórmula da derivada total (8.9) nos dá

$$\begin{aligned} \frac{dQ_a}{da} &= \frac{\partial Q_a}{\partial b} \frac{db}{da} + \frac{\partial Q_a}{\partial a} = Q_{ba} \frac{db}{da} + Q_{aa} \\ \frac{dQ_b}{da} &= \frac{\partial Q_b}{\partial b} \frac{db}{da} + \frac{\partial Q_b}{\partial a} = Q_{bb} \frac{db}{da} + Q_{ab} \end{aligned} \quad (11.43)$$

Após substituir (11.40) em (11.43) e então substituir esta última em (11.42), podemos reescrever a derivada segunda como

$$\begin{aligned} \frac{d^2b}{da^2} &= -\frac{1}{Q_b^2} \left[Q_{aa}Q_b - Q_{ba}Q_a - Q_{ab}Q_a + Q_{bb}Q_a^2 \left(\frac{1}{Q_b} \right) \right] \\ &= -\frac{1}{Q_b^3} [Q_{aa}(Q_b)^2 - 2Q_{ab}(Q_a)(Q_b) + Q_{bb}(Q_a)^2] \end{aligned} \quad (11.44)$$

Note que, em (11.44), a expressão entre colchetes (última linha) é uma forma quadrática das duas variáveis Q_a e Q_b . Se a condição suficiente de segunda ordem for satisfeita, de modo que

$$Q_{aa} < 0 \quad \text{e} \quad \begin{vmatrix} Q_{aa} & -Q_{ab} \\ -Q_{ab} & Q_{bb} \end{vmatrix} > 0$$

então, em virtude de (11.11'), a forma quadrática citada deve ser negativa definida. Isso, por sua vez, tornará d^2b/da^2 positiva, porque Q_b está restrita a ser positiva pela condição de primeira ordem. Assim, a satisfação da condição suficiente de segunda ordem significa que a isoquanta relevante (de inclinação negativa) é estritamente convexa na combinação de insumos escolhida, como afirmado.

O conceito de convexidade estrita, como aplicado a uma isoquanta $b = \phi(a)$, que é desenhada sobre um plano bidimensional ab , deve ser cuidadosamente distinguido do mesmo conceito quando aplicado à função produção $Q(a, b)$ em si, que é desenhada no espaço tridimensional abQ . Em particular, note que, se quisermos aplicar o conceito de concavidade ou convexidade estrita à função produção no presente contexto, então, para produzir a forma desejada da isoquanta, a estipulação apropriada é que $Q(a, b)$ seja estritamente *côncava* no espaço tridimensional (que tenha a forma de domo), o que contrasta claramente com a estipulação de que a isoquanta relevante seja estritamente *convexa* no espaço bidimensional (tenha forma de U totalmente ou em parte).

EXEMPLO 7

A seguir, suponha juro composto *trimestralmente* a uma taxa dada de i_0 por trimestre. Suponha, também, que o processo de produção leve exatamente um trimestre, a função lucro então se torna

$$\pi = P_0 Q(a, b)(1 + i_0)^{-1} - P_{a0}a - P_{b0}b$$

Agora, constatamos que a condição de primeira ordem é

$$P_0 Q_a(1 + i_0)^{-1} - P_{a0} = 0$$

$$P_0 Q_b(1 + i_0)^{-1} - P_{b0} = 0$$

cuja interpretação analítica é exatamente a mesma do Exemplo 6, exceto pelo modo diferente de desconto.

É fácil ver que a mesma condição suficiente obtida no Exemplo 6 também deve se aplicar aqui.

EXERCÍCIO 11.6

- Se, por outro lado, a função custo da empresa competitiva do Exemplo 1 for $C = 2Q_1^2 + 2Q_2^2$, então:
 - A produção dos dois bens ainda estará tecnicamente relacionada?
 - Quais serão os novos níveis ótimos de Q_1 e Q_2 ?
 - Qual é o valor de π_{12} ? O que isso implica em termos econômicos?
- Uma empresa de dois produtos enfrenta as seguintes funções demanda e custo:

$$Q_1 = 40 - 2P_1 - P_2 \quad Q_2 = 35 - P_1 - P_2 \quad C = Q_1^2 + 2Q_2^2 + 10$$
 - Calcule os níveis de produção que satisfaçam a condição de primeira ordem para lucro máximo. (Use frações.)
 - Verifique a condição suficiente de segunda ordem. É possível concluir que esse problema possui um único máximo absoluto?
 - Qual é o lucro máximo?
- Com base no preço e na quantidade de equilíbrio no Exemplo 4, calcule a elasticidade pontual de demanda $|\varepsilon_{dij}|$ (para $i = 1, 2$). Qual mercado tem a elasticidade de demanda mais alta e a mais baixa?
- Se a função custo do Exemplo 4 for mudada para $C = 20 + 15Q + Q_2$
 - Encontre a nova função custo marginal.
 - Encontre as novas quantidades de equilíbrio. (Use frações.)
 - Encontre os novos preços de equilíbrio
 - Verifique se a condição suficiente de segunda ordem é satisfeita.
- No Exemplo 7, como você reescreveria a função lucro se valessem as seguintes condições?
 - Os juros são compostos semestralmente a uma taxa de i_0 por ano, e o processo de produção demora um ano.
 - Os juros são compostos trimestralmente a uma taxa de i_0 por ano, e o processo de produção demora nove meses.
- Dada $Q = Q(a, b)$, como você expressaria algebricamente a isoquanta para o nível de produção de, digamos, 260?

11.7 Aspectos de estática comparativa da otimização

A otimização, que é uma variedade especial da análise estática de equilíbrio, naturalmente também está sujeita a investigações do tipo estática comparativa. Mais uma vez, a idéia é descobrir como uma variação em qualquer parâmetro afetará a posição de equilíbrio do modelo, o qual, no presente contexto, refere-se aos valores ótimos das variáveis de escolha (e ao valor ótimo da função objetivo). Uma vez que não há nenhuma técnica nova envolvida além das que discutimos na Parte 3, podemos passar diretamente a algumas ilustrações, com base nos exemplos apresentados na Seção 11.6.

Soluções de forma reduzida

O Exemplo 1 da Seção 11.6 contém dois parâmetros (ou variáveis exógenas), P_{10} e P_{20} ; portanto, não é nenhuma surpresa que os níveis ótimos de produção dessa empresa de dois produtos sejam expressos estritamente em termos desses parâmetros:

$$Q_1^* = \frac{4P_{10} - P_{20}}{15} \quad \text{e} \quad Q_2^* = \frac{4P_{20} - P_{10}}{15}$$

Essas são soluções de forma reduzida, e a simples diferenciação parcial apenas é suficiente para nos mostrar todas as propriedades estáticas comparativas do modelo, a saber,

$$\frac{\partial Q_1^*}{\partial P_{10}} = \frac{4}{15} \quad \frac{\partial Q_1^*}{\partial P_{20}} = -\frac{1}{15} \quad \frac{\partial Q_2^*}{\partial P_{10}} = -\frac{1}{15} \quad \frac{\partial Q_2^*}{\partial P_{20}} = \frac{4}{15}$$

Para lucro máximo, cada produto da empresa deve ser produzido em maior quantidade se seu preço de mercado subir ou se o preço de mercado do outro produto cair.

É claro que essas conclusões resultam somente das premissas específicas do modelo em questão. Podemos destacar, em particular, que os efeitos da variação de P_{10} sobre Q_2^* e de P_{20} sobre Q_1^* são conseqüências das relações técnicas admitidas no lado da produção dessas duas mercadorias e que, na ausência de tais relações, devemos ter

$$\frac{\partial Q_1^*}{\partial P_{20}} = \frac{\partial Q_2^*}{\partial P_{10}} = 0$$

Passando para o Exemplo 2, notamos que os níveis ótimos de produção estão ali estabelecidos, numericamente, como $Q_1^* = 8$ e $Q_2^* = 7\frac{2}{3}$ – nenhum parâmetro aparece. De fato, todas as constantes nas equações do modelo são numéricas em vez de paramétricas, de modo que, quando atingirmos o estágio de solução, essas constantes terão perdido suas respectivas identidades por meio do processo de manipulação aritmética. Isso serve para acentuar a falta fundamental de generalidade na utilização de constantes numéricas e a conseqüente falta de conteúdo comparativo estático na solução de equilíbrio.

Por outro lado, a não-utilização de constantes numéricas não garante que um problema se tornará automaticamente passível de análise estática comparativa. O problema da discriminação de preços (Exemplo 3), por exemplo, foi enunciado primordialmente para o estudo da condição de equilíbrio (maximização de lucro) e absolutamente nenhum parâmetro foi introduzido. De acordo com isso, mesmo que enunciado em termos de funções gerais, será necessária uma reformulação se quisermos fazer um estudo de estática comparativa.

Modelos de função geral

O problema da decisão de insumos do Exemplo 6 ilustra o caso em que a formulação de uma função geral realmente abrange diversos parâmetros – de fato, não menos que cinco (P_0 , P_{a0} , P_{b0} , r e t), nos quais omitimos, como antes, os índices 0 das variáveis exógenas r_0 e t_0 . Como obtemos as propriedades estáticas comparativas desse modelo?

Mais uma vez, a resposta está na aplicação do teorema da função implícita. Porém, diferentemente dos casos de modelos de equilíbrio de mercado ou de determinação de renda nacional sem metas definidas, em que trabalhamos com as condições de equilíbrio do modelo, o presente contexto de equilíbrio com metas definidas determina que trabalhem com as condições de otimização de primeira ordem. Para o Exemplo 6, essas condições são enunciadas em (11.39). Reunindo todos os termos em (11.39) à esquerda do sinal de igualdade, e explicitando que Q_a e Q_b são, ambas, funções das variáveis endógenas (de escolha) a e b , podemos reescrever as condições de primeira ordem no formato de (8.24), como se segue:

$$\begin{aligned} F^1(a, b; P_0, P_{a0}, P_{b0}, r, t) &= P_0 Q_a(a, b) e^{-rt} - P_{a0} = 0 \\ F^2(a, b; P_0, P_{a0}, P_{b0}, r, t) &= P_0 Q_b(a, b) e^{-rt} - P_{b0} = 0 \end{aligned} \quad (11.45)$$

Supomos que as funções F^1 e F^2 possuem derivadas contínuas. Assim, seria possível aplicar o teorema da função implícita, contanto que o jacobiano desse sistema em relação às variáveis endógenas a e b não se anule no equilíbrio inicial. O citado jacobiano revela ser nada mais que o determinante hessiano da função π do Exemplo 6:

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial a} & \frac{\partial F^1}{\partial b} \\ \frac{\partial F^2}{\partial a} & \frac{\partial F^2}{\partial b} \end{vmatrix} = \begin{vmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{vmatrix} = |H| \text{ [por (11.41)]} \quad (11.46)$$

Portanto, se supusermos que a condição suficiente de segunda ordem para a maximização do lucro está satisfeita, então $|H|$ deve ser positivo, assim como $|J|$, no equilíbrio inicial ou no equilíbrio ótimo. Nesse caso, o teorema da função implícita nos habilitará a escrever o par de funções implícitas

$$\begin{aligned} a^* &= a^*(P_0, P_{a0}, P_{b0}, r, t) \\ b^* &= b^*(P_0, P_{a0}, P_{b0}, r, t) \end{aligned} \quad (11.47)$$

bem como o par de identidades

$$\begin{aligned} P_0 Q_a(a^*, b^*) e^{-rt} - P_{a0} &\equiv 0 \\ P_0 Q_b(a^*, b^*) e^{-rt} - P_{b0} &\equiv 0 \end{aligned} \quad (11.48)$$

Para estudar a estática comparativa do modelo, em primeiro lugar, tome a diferencial total de cada identidade em (11.48). Por enquanto, permitiremos que todas as variáveis exógenas variem, de modo que o resultado da diferenciação total envolverá da^* , db^* , bem como dP_0 , dP_{a0} , dP_{b0} , dr e dt . Se colocarmos do lado esquerdo do sinal de igualdade apenas os termos que envolvem da^* e db^* , o resultado será

$$\begin{aligned} P_0 Q_{aa} e^{-rt} da^* + P_0 Q_{ab} e^{-rt} db^* &= -Q_a e^{-rt} dP_0 + dP_{a0} \\ &\quad + P_0 Q_a t e^{-rt} dr + P_0 Q_a r e^{-rt} dt \\ P_0 Q_{ab} e^{-rt} da^* + P_0 Q_{bb} e^{-rt} db^* &= -Q_b e^{-rt} dP_0 + dP_{b0} \\ &\quad + P_0 Q_b t e^{-rt} dr + P_0 Q_b r e^{-rt} dt \end{aligned} \quad (11.49)$$

onde, note bem, as derivadas primeira e segunda de Q devem ser todas calculadas no ponto de equilíbrio, isto é, em a^* e b^* . Você notará, também, que os coeficientes de da^* e db^* à esquerda são exatamente os elementos do jacobiano em (11.46).

Para obter as derivadas estáticas comparativas específicas – das quais há um total de 10 (por quê?) – agora permitiremos que somente uma única variável exógena varie por vez. Suponha

que permitimos que apenas P_0 varie. Então, $dP_0 \neq 0$, mas $dP_{a0} = dP_{b0} = dr = dt = 0$, de modo que restará somente o primeiro termo no lado direito de cada equação em (11.49). Dividindo tudo por dP_0 , e interpretando a razão da^*/dP_0 como a derivada estática comparativa $(\partial a^*/\partial P_0)$ e a razão db^*/dP_0 de modo semelhante, podemos escrever a equação matricial

$$\begin{bmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial a^*/\partial P_0) \\ (\partial b^*/\partial P_0) \end{bmatrix} = \begin{bmatrix} -Q_a e^{-rt} \\ -Q_b e^{-rt} \end{bmatrix}$$

Constatamos que a solução, pela regra de Cramer é

$$\begin{aligned} \left(\frac{\partial a^*}{\partial P_0} \right) &= \frac{(Q_b Q_{ab} - Q_a Q_{bb}) P_0 e^{-2rt}}{|J|} \\ \left(\frac{\partial b^*}{\partial P_0} \right) &= \frac{(Q_a Q_{ab} - Q_b Q_{aa}) P_0 e^{-2rt}}{|J|} \end{aligned} \quad (11.50)$$

Se preferir, há um método alternativo para obter esses resultados: basta simplesmente diferenciar *totalmente* as duas identidades em (11.48) em relação a P_0 (mantendo fixas as outras quatro variáveis exógenas), sem esquecer que P_0 pode afetar a^* e b^* por meio de (11.47).

Agora vamos analisar os sinais das derivadas estáticas comparativas em (11.50). Sob a hipótese de que a condição de segunda ordem está satisfeita, o jacobiano no denominador deve ser positivo. A condição de segunda ordem também implica que Q_{aa} e Q_{bb} são negativas, exatamente como a condição de primeira ordem implica que Q_a e Q_b são positivas. Além disso, a expressão $P_0 e^{-2rt}$ é, com certeza, positiva. Assim, se $Q_{ab} > 0$ (se o crescimento de um insumo elevar o MPP do outro insumo), podemos concluir que ambas, $(\partial a^*/\partial P_0)$ e $(\partial b^*/\partial P_0)$, serão positivas, implicando que um aumento no preço do produto resultará em maior emprego de ambos os insumos no equilíbrio. Se, por outro lado, $Q_{ab} < 0$, o sinal de cada derivada em (11.50) dependerá das intensidades relativas da força negativa e da força positiva na expressão entre parênteses à direita.

Em seguida, vamos permitir que apenas a variável exógena r varie. Então, todos os termos à direita de (11.49) se anularão, exceto os que envolvem dr . Dividindo tudo por $dr \neq 0$, obtemos agora a seguinte equação matricial

$$\begin{bmatrix} P_0 Q_{aa} e^{-rt} & P_0 Q_{ab} e^{-rt} \\ P_0 Q_{ab} e^{-rt} & P_0 Q_{bb} e^{-rt} \end{bmatrix} \begin{bmatrix} (\partial a^*/\partial r) \\ (\partial b^*/\partial r) \end{bmatrix} = \begin{bmatrix} P_0 Q_a t e^{-rt} \\ P_0 Q_b t e^{-rt} \end{bmatrix}$$

cujas soluções são

$$\begin{aligned} \left(\frac{\partial a^*}{\partial r} \right) &= \frac{t(Q_a Q_{bb} - Q_b Q_{ab}) (P_0 e^{-rt})^2}{|J|} \\ \left(\frac{\partial b^*}{\partial r} \right) &= \frac{t(Q_b Q_{aa} - Q_a Q_{ab}) (P_0 e^{-rt})^2}{|J|} \end{aligned} \quad (11.51)$$

Essas duas derivadas estáticas comparativas serão ambas negativas se Q_{ab} for positiva, mas seus sinais serão indeterminados se Q_{ab} for negativa.

Por um procedimento semelhante, podemos descobrir os efeitos das variações nos parâmetros restantes. Na verdade, em vista da simetria entre r e t em (11.48), fica imediatamente óbvio que a aparência de ambas, $(\partial a^*/\partial t)$ e $(\partial b^*/\partial t)$, deve ser semelhante a (11.51).

Deixamos que você analise os efeitos das variações em P_{a0} e P_{b0} . Como você constatará, a restrição de sinal da condição suficiente de segunda ordem novamente será útil para calcular as derivadas estáticas comparativas, porque ela pode nos informar os sinais de Q_{aa} e Q_{bb} , bem como o jacobiano $|J|$ no equilíbrio inicial (ótimo). Assim, à parte distinguir entre máximo e mínimo, a condição de segunda ordem também tem um papel vital a desempenhar no estudo dos deslocamentos de posições de equilíbrio.

EXERCÍCIO 11.7

Para os Problemas de 1 a 3, suponha que $Q_{ab} > 0$.

1. Tendo como base o modelo descrito em (11.45) até (11.48), encontre as derivadas estáticas comparativas $(\partial a^* / \partial P_{a0})$ e $(\partial b^* / \partial P_{a0})$. Interprete o significado econômico do resultado. Em seguida, analise os efeitos de uma variação de P_{b0} sobre a^* e b^* .
2. Para o problema do Exemplo 7 na Seção 11.6:
 - (a) Há quantos parâmetros? Enumere-os.
 - (b) Seguindo o procedimento descrito em (11.45) até (11.50) e supondo que a condição suficiente de segunda ordem está satisfeita, encontre as derivadas estáticas comparativas $(\partial a^* / \partial P_0)$ e $(\partial b^* / \partial P_0)$. Calcule seus sinais e interprete seus significados econômicos.
 - (c) Encontre $(\partial a^* / \partial i_0)$ e $(\partial b^* / \partial i_0)$, calcule seus sinais e interprete seus significados econômicos.
3. Mostre que os resultados em (11.50) podem ser obtidos de modo alternativo diferenciando *totalmente* as duas identidades em (11.48) em relação a P_0 , mantendo fixas as outras variáveis exógenas. Não se esqueça que P_0 pode afetar a^* e b^* em virtude de (11.47).
4. Um determinante jacobiano, como definido em (7.27), é composto de derivadas parciais de *primeira* ordem. Por outro lado, os elementos de um determinante hessiano, como definido nas Seções 11.3 e 11.4, são derivadas parciais de *segunda ordem*. Então, como é que $|J| = |H|$, como em (11.46)?

CAPÍTULO 12

Otimização com restrições de igualdade

O Capítulo 11 apresentou um método geral para achar os extremos relativos de uma função objetivo de duas ou mais variáveis de escolha. Um aspecto importante daquela discussão é que todas as variáveis de escolha são *independentes* umas das outras, no sentido de que a decisão referente a uma variável não influi na escolha das variáveis restantes. Por exemplo, uma empresa de dois produtos pode escolher qualquer valor para Q_1 e qualquer valor para Q_2 que desejar, sem que uma escolha limite a outra.

Se a citada empresa for obrigada a observar uma restrição (por exemplo, uma quota de produção) na forma de $Q_1 + Q_2 = 950$, entretanto, a independência entre as variáveis de escolha deixará de existir. Nesse caso, os níveis de produto Q_1^* e Q_2^* que maximizam o lucro da empresa não serão apenas simultâneos, mas também dependentes porque, quanto mais alto for Q_1^* , mais baixo Q_2^* deverá ser, proporcionalmente, de modo a ficar dentro da quota combinada de 950. O novo ótimo que satisfaz a quota de produção constitui um *ótimo restrito* que, em geral, poderá ser diferente do *ótimo livre* discutido no Capítulo 11.

Uma restrição, tal como a quota de produção mencionada anteriormente, estabelece uma relação entre as duas variáveis no que se refere a seus papéis como variáveis de escolha, mas essa relação deve ser distinta de outros tipos de relação que possam ligar as variáveis entre si. Por exemplo, no Exemplo 2 da Seção 11.6, os dois produtos da empresa estão relacionados pelo consumo (são substitutos), bem como pela produção (o que se reflete na função custo), mas esse fato não qualifica o problema como de otimização restrita, já que as duas variáveis de produção ainda são *independentes enquanto variáveis de escolha*. Somente a dependência das variáveis enquanto variáveis de escolha dá origem a um ótimo restrito.

Neste capítulo, consideraremos apenas restrições de igualdade, tal como $Q_1 + Q_2 = 950$. Nossa preocupação principal serão os extremos restritos *relativos*, embora os *absolutos* também sejam discutidos na Seção 12.4.

12.1 Efeitos de uma restrição

O objetivo principal de impor uma restrição é reconhecer devidamente certos fatores limitantes presentes no problema em discussão.

Já vimos a limitação sobre as escolhas de níveis de produto que resultam de uma quota de produção. Para ilustrar melhor, vamos considerar um consumidor cuja função utilidade simples (índice) é

$$U = x_1 x_2 + 2x_1 \quad (12.1)$$

Visto que aqui as utilidades marginais – as derivadas parciais $U_1 \equiv \partial U / \partial x_1$ e $U_2 \equiv \partial U / \partial x_2$ – são positivas para todos os níveis positivos de x_1 e x_2 , maximizar U sem nenhuma restrição significa que o consumidor deve comprar uma quantidade *infinita* de ambos os bens, uma solução que obviamente tem pouca relevância prática. Para tornar o problema de otimização significativo, o poder de compra do consumidor também deve ser levado em consideração, isto é, deve ser incorporada ao problema uma *restrição orçamentária*. Se o consumidor pretende gastar uma determinada quantia, digamos, \$60, nos dois bens, e se os preços correntes são $P_{10} = 4$ e $P_{20} = 2$, então a restrição orçamentária pode ser expressa pela equação linear

$$4x_1 + 2x_2 = 60 \quad (12.2)$$

Tal restrição, assim como a quota de produção a que nos referimos antes, resulta na escolha de x_1^* e x_2^* mutuamente dependentes.

O problema agora é maximizar (12.1), sujeita à restrição declarada em (12.2). Matematicamente, o que a restrição (também denominada *vínculo*, *relação lateral* ou *condição subsidiária*) faz é estreitar o domínio e, por conseguinte, a imagem da função objetivo. O domínio de (12.1) normalmente seria o conjunto $\{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0\}$. A representação gráfica desse domínio é o quadrante não-negativo do plano $x_1 x_2$ na Figura 12.1a. Após a adição da restrição orçamentária (12.2), entretanto, podemos admitir apenas os valores das variáveis que satisfaçam esta última equação, de modo que o domínio é imediatamente reduzido ao conjunto de pontos que estão sobre a linha de orçamento. Isso também afetará automaticamente a imagem da função objetivo; somente o subconjunto da superfície de utilidade que está diretamente acima da linha da restrição orçamentária será relevante agora. O subconjunto citado (uma seção transversal da superfície) pode ser parecido com a curva na Figura 12.1b, onde os valores de U estão sobre o eixo vertical e a linha de orçamento do diagrama *a* está sobre o eixo horizontal. Então, nosso interesse é apenas localizar o máximo na curva do diagrama *b*.

Em geral, para uma função $z = f(x, y)$, a diferença entre um extremo restrito e um extremo livre pode ser ilustrada no gráfico tridimensional da Figura 12.2. O extremo livre nesse gráfico específico é o ponto correspondente ao pico do domo inteiro, mas o extremo restrito está no ponto de pico da curva em forma de U invertido construída sobre a reta de restrição (isto é, diretamente acima dela). Em geral, pode-se esperar que o valor de um máximo restrito seja mais baixo do que o de um máximo livre, embora, por coincidência, pode acontecer de os dois máximos terem o mesmo valor. Mas o máximo restrito nunca pode exceder o máximo livre.

É interessante notar que, se tivéssemos adicionado uma outra restrição que interceptasse a primeira restrição em um único ponto no plano xy , as duas restrições juntas teriam restringido o domínio àquele único ponto. Então, a localização do extremo se tornaria uma questão trivial. Quando se trata de um problema significativo, o número e a natureza dos vínculos devem ser tais que restrinjam, mas não eliminem, a possibilidade de escolha. Em geral, o número de restrições deve ser menor que o número de variáveis de escolha.

FIGURA 12.1

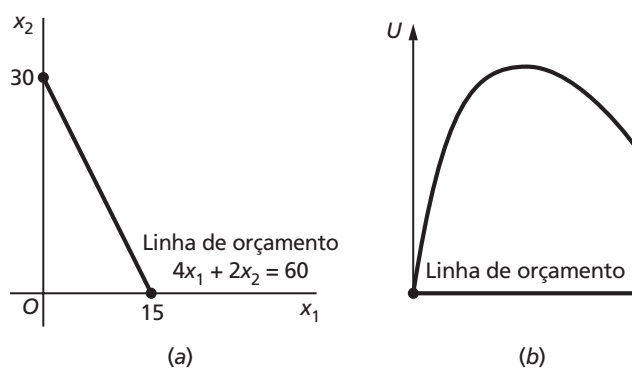
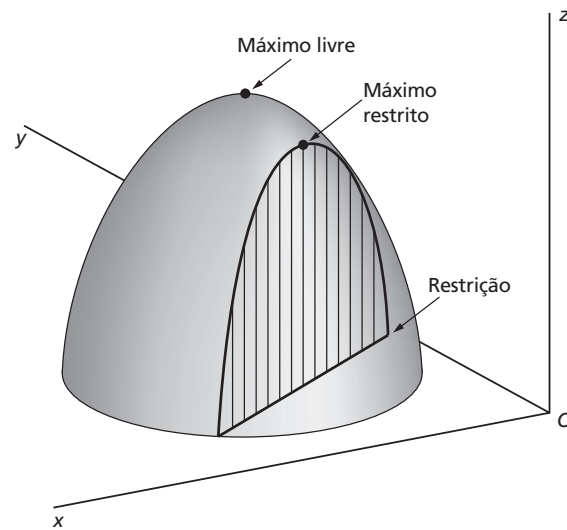


FIGURA 12.2



12.2 Achando os valores estacionários

Mesmo sem nenhuma nova técnica de solução, o máximo restrito no exemplo simples definido por (12.1) e (12.2) pode ser encontrado com facilidade. Uma vez que a restrição (12.2) implica

$$x_2 = \frac{60 - 4x_1}{2} = 30 - 2x_1 \quad (12.2')$$

podemos combinar a restrição com a função objetivo substituindo (12.2') em (12.1). O resultado é uma função objetivo com somente uma variável:

$$U = x_1(30 - 2x_1) + 2x_1^2 = 32x_1 - 2x_1^2$$

que pode ser manipulada pelo método que já aprendemos. Igualando $dU/dx_1 = 32 - 4x_1$ a zero, obtemos a solução $x_1^* = 8$, a qual, em virtude de (12.2'), leva imediatamente a $x_2^* = 30 - 2(8) = 14$. Usando (12.1), podemos achar o valor estacionário $U^* = 128$; e, visto que a derivada segunda é $d^2U/dx_1^2 = -4 < 0$, aquele valor estacionário constitui um máximo (restrito) de U .[†]

Contudo, quando a restrição é, em si, uma função complicada, ou quando há várias restrições a considerar, a técnica de substituição e eliminação de variáveis pode tornar-se uma tarefa trabalhosa. E, mais importante, quando a restrição é apresentada sob uma forma tal que não podemos resolvê-la para expressar uma variável (x_2) como uma função explícita da outra (x_1), o método de eliminação seria, de fato, em vão – ainda que soubéssemos de antemão que x_2 é uma função implícita de x_1 , isto é, mesmo que as condições do teorema da função implícita fossem satisfeitas. Nesses casos, podemos recorrer a um método conhecido como o *método do multiplicador (indeterminado) de Lagrange* que, como veremos, tem vantagens analíticas específicas.

Método do multiplicador de Lagrange

A essência do método do multiplicador de Lagrange é converter um problema de extremo restrito em uma forma tal que a condição de primeira ordem do problema do extremo livre ainda possa ser aplicada.

[†] Recorde que, no problema do canteiro de flores do Exercício 9.4-2, foi aplicada a mesma técnica de substituição para achar a área máxima, utilizando uma restrição (a quantidade disponível de alambrado) para eliminar uma das variáveis (o comprimento ou a largura do canteiro).

Dado o problema de maximizar $U = x_1 x_2 + 2x_1$, sujeita à restrição $4x_1 + 2x_2 = 60$ [de (12.1) e (12.2)], vamos escrever a denominada *função de Lagrange*, que é uma versão modificada da função objetivo que incorpora a restrição como se segue:

$$Z = x_1 x_2 + 2x_1 + \lambda(60 - 4x_1 - 2x_2) \quad (12.3)$$

O símbolo λ (a letra grega lambda), que representa algum número ainda não determinado, é denominado um *multiplicador (indeterminado) de Lagrange*. Se de algum modo pudermos ter certeza de que $4x_1 + 2x_2 = 60$, de modo que a restrição seja satisfeita, então o último termo em (12.3) se anulará independentemente do valor de λ . Nesse caso, Z será idêntica a U . Além do mais, sem a restrição para atrapalhar, basta procurar o máximo *livre* de Z , em lugar do *máximo restrito* de U , em relação às duas variáveis x_1 e x_2 . A questão é: como anular a expressão entre parênteses em (12.3)?

A tática que conseguirá tal feito é simplesmente tratar λ como uma variável de escolha adicional em (12.3), isto é, considerar $Z = Z(\lambda, x_1, x_2)$. Pois, então, a condição de primeira ordem para um extremo livre consistirá em um conjunto de equações simultâneas

$$\begin{aligned} Z_\lambda (\equiv \partial Z / \partial \lambda) &= 60 - 4x_1 - 2x_2 = 0 \\ Z_1 (\equiv \partial Z / \partial x_1) &= x_2 + 2 - 4\lambda = 0 \\ Z_2 (\equiv \partial Z / \partial x_2) &= x_1 - 2\lambda = 0 \end{aligned} \quad (12.4)$$

e a primeira equação garantirá automaticamente a satisfação da restrição. Assim, incorporando a restrição à função de Lagrange Z , e tratando o multiplicador de Lagrange como uma variável extra, podemos obter o extremo restrito U^* (duas variáveis de escolha) simplesmente filtrando os valores estacionários de Z , tomada como uma função *livre* de três variáveis de escolha.

Resolvendo (12.4) para os valores críticos das variáveis, encontramos $x_1^* = 8$, $x_2^* = 14$ (e $\lambda^* = 4$). Como esperado, os valores de x_1^* e x_2^* estão de acordo com as respostas já obtidas pelo método de substituição. Além do mais, fica claro, a partir de (12.3), que $Z^* = 128$; esse valor é, como deve ser, idêntico ao valor de U^* encontrado antes.

Em geral, dada uma função objetivo

$$z = f(x, y) \quad (12.5)$$

sujeita à restrição

$$g(x, y) = c \quad (12.6)$$

onde c é uma constante,[†] podemos escrever a função de Lagrange como

$$Z = f(x, y) + \lambda[c - g(x, y)] \quad (12.7)$$

Para valores estacionários de Z , considerada uma função das três variáveis λ , x e y , a condição necessária é

$$\begin{aligned} Z_\lambda &= c - g(x, y) = 0 \\ Z_x &= f_x - \lambda g_x = 0 \\ Z_y &= f_y - \lambda g_y = 0 \end{aligned} \quad (12.8)$$

Uma vez que a primeira equação em (12.8) é simplesmente uma nova maneira de enunciar (12.6), os valores estacionários da função de Lagrange Z automaticamente satisfarão a restrição

[†] Também é possível incorporar a constante c à função de restrição de modo que (12.6) seja escrita alternativamente como $G(x, y) = 0$, onde $G(x, y) = g(x, y) - c$. Nesse caso, (12.7) deve ser mudada para $Z = f(x, y) + \lambda[0 - G(x, y)] = f(x, y) - \lambda G(x, y)$. A versão em (12.6) é escolhida porque facilita o estudo do efeito estático comparativo de uma variação ulterior na constante de restrição [veja (12.16)].

da função original z . E, visto que a expressão $\lambda[c - g(x, y)]$ agora é, com certeza, igual a zero, os valores estacionários de Z em (12.7) devem ser idênticos aos de (12.5), sujeita a (12.6).

Vamos ilustrar o método com mais dois exemplos.

EXEMPLO 1 Calcule o extremo de

$$z = xy \quad \text{sujeita a} \quad x + y = 6$$

O primeiro passo é escrever a função de Lagrange

$$Z = xy + \lambda(6 - x - y)$$

Para um valor estacionário de Z , é necessário que

$$\left. \begin{aligned} Z_\lambda &= 6 - x - y = 0 \\ Z_x &= y - \lambda = 0 \\ Z_y &= x - \lambda = 0 \end{aligned} \right\} \text{ ou } \begin{cases} x + y = 6 \\ -\lambda + y = 0 \\ -\lambda + x = 0 \end{cases}$$

Assim, pela regra de Cramer ou por qualquer outro método, podemos chegar a

$$\lambda^* = 3 \quad x^* = 3 \quad y^* = 3$$

O valor estacionário é $Z^* = z^* = 9$, que precisa ser testado em relação à condição de segunda ordem antes de podermos dizer se ele é um máximo ou um mínimo (ou nenhum dos dois). Trataremos disso na Seção 12.3.

EXEMPLO 2 Calcule o extremo de

$$z = x_1^2 + x_2^2 \quad \text{sujeita a} \quad x_1 + 4x_2 = 2$$

A função de Lagrange é

$$Z = x_1^2 + x_2^2 + \lambda(2 - x_1 - 4x_2)$$

para a qual a condição necessária para um valor estacionário é

$$\left. \begin{aligned} Z_\lambda &= 2 - x_1 - 4x_2 = 0 \\ Z_1 &= 2x_1 - \lambda = 0 \\ Z_2 &= 2x_2 - 4\lambda = 0 \end{aligned} \right\} \text{ ou } \begin{cases} x_1 + 4x_2 = 2 \\ -\lambda + 2x_1 = 0 \\ -4\lambda + 2x_2 = 0 \end{cases}$$

O valor estacionário de Z , definido pela solução

$$\lambda^* = \frac{4}{17} \quad x_1^* = \frac{2}{17} \quad x_2^* = \frac{8}{17}$$

é, portanto, $Z^* = z^* = \frac{4}{17}$. Novamente, uma condição de segunda ordem deve ser consultada antes de podermos dizer se z^* é um máximo ou um mínimo.

Abordagem da diferencial total

Quando discutimos o extremo livre de $z = f(x, y)$, aprendemos que a condição necessária de primeira ordem pode ser enunciada em termos da diferencial total dz como se segue:

$$dz = f_x dx + f_y dy = 0 \quad (12.9)$$

Essa afirmação permanece válida após a adição de uma restrição $g(x, y) = c$. Contudo, na presença da restrição, não podemos mais considerar ambas, dx e dy , como variações “arbitrárias”, como antes. Pois, se $g(x, y) = c$, então dg deve ser igual a dc , a qual é zero, já que c é uma constante. Portanto

$$(dg =) g_x dx + g_y dy = 0 \quad (12.10)$$

e essa relação torna dx e dy dependentes uma da outra. Por conseguinte, a condição necessária de primeira ordem se torna $dz = 0$ [(12.9)], sujeita a $g = c$, e, por conseqüência, também sujeita a $dg = 0$ [(12.10)]. Uma inspeção visual de (12.9) e (12.10) deve deixar claro que, para satisfazer essa condição necessária, devemos ter

$$\frac{f_x}{g_x} = \frac{f_y}{g_y} \quad (12.11)$$

Esse resultado pode ser verificado resolvendo (12.10) para dy e substituindo o resultado em (12.9). A condição (12.11), juntamente com a restrição $g(x, y) = c$, fornecerá duas equações a partir das quais calcularemos os valores críticos de x e y .[†]

A abordagem da diferencial total resulta na mesma condição de primeira ordem do método do multiplicador de Lagrange? Vamos comparar (12.8) com o resultado que acabamos de obter. A primeira equação em (12.8) é uma mera repetição da restrição; o novo resultado requer que ela também seja satisfeita. As duas últimas equações em (12.8) podem ser reescritas, respectivamente, como

$$\frac{f_x}{g_x} = \lambda \quad \text{e} \quad \frac{f_y}{g_y} = \lambda \quad (12.11')$$

e elas transmitem exatamente as mesmas informações que (12.11). Todavia, note que, conquanto a abordagem da diferenciação total resulte somente nos valores de x^* e y^* , o método do multiplicador de Lagrange também dá o valor de λ^* como um subproduto direto. Acontece que λ^* nos dá uma medida da sensibilidade de Z^* (e z^*) a uma mudança na restrição, como demonstraremos em breve. Portanto, o método do multiplicador de Lagrange oferece a vantagem de conter certas informações de estática comparativa embutidas na solução.

Uma interpretação do multiplicador de Lagrange

Para mostrar que λ^* realmente mede a sensibilidade de Z^* a variações na restrição, vamos executar uma análise estática comparativa da condição de primeira ordem (12.8). Visto que λ , x e y são endógenas, a única variável exógena disponível é o parâmetro de restrição c . Uma variação em c causaria um deslocamento da curva de restrição no plano xy e, por conseguinte, alteraria a solução ótima. Em particular, o efeito de um *aumento* em c (um orçamento maior ou uma quota de produção maior) indicaria como a solução ótima é afetada por um *abrandamento* da restrição.

Para fazer a análise estática comparativa, recorreremos novamente ao teorema da função implícita. Supondo que as três equações em (12.8) estão na forma $F^j(\lambda, x, y; c) = 0$ (com $j = 1, 2, 3$) e supondo que elas têm derivadas parciais contínuas, devemos primeiro verificar se o seguinte jacobiano de variáveis endógenas (no qual $f_{xy} = f_{yx}$ e $g_{xy} = g_{yx}$)

[†] Note que a restrição $g = c$ ainda tem de ser considerada juntamente com (12.11), ainda que tenhamos utilizado a equação $dg = 0$ – isto é, (12.10) – para obter (12.11). Enquanto $g = c$ implica necessariamente $dg = 0$, a recíproca não é verdadeira: $dg = 0$ implica apenas que g = uma constante (não necessariamente c). Por conseguinte, a menos que a restrição seja explicitamente considerada, alguma informação ficará inadvertidamente de fora do problema.

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial \lambda} & \frac{\partial F^1}{\partial x} & \frac{\partial F^1}{\partial y} \\ \frac{\partial F^2}{\partial \lambda} & \frac{\partial F^2}{\partial x} & \frac{\partial F^2}{\partial y} \\ \frac{\partial F^3}{\partial \lambda} & \frac{\partial F^3}{\partial x} & \frac{\partial F^3}{\partial y} \end{vmatrix} = \begin{vmatrix} 0 & -g_x & -g_y \\ -g_x & f_{xx} - \lambda g_{xx} & f_{xy} - \lambda g_{xy} \\ -g_y & f_{xy} - \lambda g_{xy} & f_{yy} - \lambda g_{yy} \end{vmatrix} \quad (12.12)$$

não se anula no estado ótimo. Nesse momento, certamente não há nenhuma indicação de que esse seria o caso. Mas nossa experiência anterior com a estática comparativa de problemas de otimização [veja a discussão de (11.46)] sugeriria que esse jacobiano está estreitamente relacionado com a condição suficiente de segunda ordem e que, se a condição suficiente for satisfeita, então o jacobiano será diferente de zero no equilíbrio (ótimo). Vamos deixar a demonstração completa desse fato para a Seção 12.3 e prosseguir supondo que $|J| \neq 0$. Isso posto, então podemos expressar λ^* , x^* e y^* como funções implícitas do parâmetro c :

$$\lambda^* = \lambda^*(c) \quad x^* = x^*(c) \quad \text{e} \quad y^* = y^*(c) \quad (12.13)$$

todas as quais terão derivadas contínuas. Além disso, temos as identidades de equilíbrio

$$\begin{aligned} c - g(x^*, y^*) &\equiv 0 \\ f_x(x^*, y^*) - \lambda^* g_x(x^*, y^*) &\equiv 0 \\ f_y(x^*, y^*) - \lambda^* g_y(x^*, y^*) &\equiv 0 \end{aligned} \quad (12.14)$$

Agora, visto que o valor ótimo de Z depende de λ^* , x^* e y^* , isto é,

$$Z^* = f(x^*, y^*) + \lambda^*[c - g(x^*, y^*)] \quad (12.15)$$

podemos, em vista de (12.13), considerar Z^* uma função apenas de c . Diferenciando totalmente Z^* em relação a c , encontramos

$$\begin{aligned} \frac{dZ^*}{dc} &= f_x \frac{dx^*}{dc} + f_y \frac{dy^*}{dc} + [c - g(x^*, y^*)] \frac{d\lambda^*}{dc} + \lambda^* \left(1 - g_x \frac{dx^*}{dc} - g_y \frac{dy^*}{dc} \right) \\ &= (f_x - \lambda^* g_x) \frac{dx^*}{dc} + (f_y - \lambda^* g_y) \frac{dy^*}{dc} + [c - g(x^*, y^*)] \frac{d\lambda^*}{dc} + \lambda^* \end{aligned}$$

onde f_x , f_y , g_x e g_y devem ser todas avaliadas no ótimo. Contudo, por (12.14), os três primeiros termos da direita serão descartados. Assim, ficamos com o resultado simples

$$\frac{dZ^*}{dc} = \lambda^* \quad (12.16)$$

que valida o que afirmamos, isto é, que o valor de solução do multiplicador de Lagrange constitui uma medida do efeito de uma variação na restrição por meio do parâmetro c sobre o valor ótimo da função objetivo.

Porém, aqui talvez caiba um alerta. Para essa interpretação de λ^* , você deve expressar Z especificamente como em (12.7). Em particular, escreva o último termo como $\lambda[c - g(x, y)]$, e não $\lambda[g(x, y) - c]$.

Casos de n variáveis e múltiplas restrições

A generalização do método do multiplicador de Lagrange para n variáveis pode ser feita com facilidade se escrevermos as variáveis de escolha usando a notação com índices. Então, a função objetivo terá a forma

$$z = f(x_1, x_2, \dots, x_n)$$

sujeita à restrição

$$g(x_1, x_2, \dots, x_n) = c$$

Segue-se que a função de Lagrange será

$$Z = f(x_1, x_2, \dots, x_n) + \lambda[c - g(x_1, x_2, \dots, x_n)]$$

para a qual a condição de primeira ordem consistirá nas seguintes $(n+1)$ equações simultâneas:

$$Z_\lambda = c - g(x_1, x_2, \dots, x_n) = 0$$

$$Z_1 = f_1 - \lambda g_1 = 0$$

$$Z_2 = f_2 - \lambda g_2 = 0$$

$$Z_n = f_n - \lambda g_n = 0$$

Mais uma vez, a primeira dessas equações nos garantirá que a restrição seja cumprida, ainda que devamos focalizar nossa atenção na função de Lagrange *livre*.

Quando há mais de uma restrição, o método do multiplicador de Lagrange é igualmente aplicável, contanto que sejam introduzidos tantos desses multiplicadores quantos forem as restrições na função de Lagrange. Suponha que uma função de n variáveis está sujeita simultaneamente às duas restrições

$$g(x_1, x_2, \dots, x_n) = c \quad \text{e} \quad h(x_1, x_2, \dots, x_n) = d$$

Então, adotando λ e μ (a letra grega mu) como os dois multiplicadores indeterminados, podemos construir uma função de Lagrange como se segue:

$$Z = f(x_1, x_2, \dots, x_n) + \lambda[c - g(x_1, x_2, \dots, x_n)] + \mu[d - h(x_1, x_2, \dots, x_n)]$$

Essa função terá o mesmo valor que a função objetivo original f se ambas as restrições forem satisfeitas, isto é, se os dois últimos termos da função de Lagrange se anularem. Considerando λ e μ como variáveis de escolha, agora contamos $(n+2)$ variáveis e, assim, a condição de primeira ordem consistirá, nesse caso, nas $(n+2)$ equações simultâneas seguintes:

$$Z_\lambda = c - g(x_1, x_2, \dots, x_n) = 0$$

$$Z_\mu = d - h(x_1, x_2, \dots, x_n) = 0$$

$$Z_i = f_i - \lambda g_i - \mu h_i = 0 \quad (i = 1, 2, \dots, n)$$

Essas equações normalmente nos habilitariam a resolver para todos os x_i , bem como para λ e μ . Como antes, as duas primeiras equações da condição necessária representam, em essência, simplesmente um outro enunciado das duas restrições.

EXERCÍCIO 12.2

- Use o método do multiplicador de Lagrange para calcular os valores estacionários de z :
 - $z = xy$, sujeita a $x + 2y = 2$.
 - $z = x(y + 4)$, sujeita a $x + y = 8$.
 - $z = x - 3y - xy$, sujeita a $x + y = 6$.
 - $z = 7 - y + x^2$, sujeita a $x + y = 0$.
- No Problema 1, verifique se um ligeiro abrandamento da restrição aumentaria ou reduziria o valor ótimo de z . A que taxa?
- Escreva a função de Lagrange e a condição de primeira ordem para valores estacionários (sem resolver as equações) para cada uma das seguintes:
 - $z = x + 2y + 3w + xy - yw$, sujeita a $x + y + 2w = 10$.
 - $z = x^2 + 2xy + yw^2$, sujeita a $2x + y + w^2 = 24$ e $x + w = 8$.
- Se, em vez de $g(x, y) = c$, a restrição fosse escrita na forma $G(x, y) = 0$, como, em consequência, seriam modificadas a função de Lagrange e a condição de primeira ordem?
- Quando discutimos a abordagem da diferencial total, salientamos que, dada a restrição $g(x, y) = c$, podemos deduzir que $dg = 0$. Pelo mesmo critério, podemos deduzir ainda mais que $d^2g = d(dg) = d(0) = 0$. Ainda assim, quando da nossa discussão anterior do extremo sem restrição de uma função $z = f(x, y)$, tivemos a situação em que $dz = 0$ era acompanhada de uma d^2z positiva definida ou negativa definida, em vez de $d^2z = 0$. Como você explicaria essa disparidade no tratamento dos dois casos?
- Se a função de Lagrange for escrita como $Z = f(x, y) + \lambda[g(x, y) - c]$, e não como em (12.7), ainda podemos interpretar o multiplicador de Lagrange como em (12.16)? Dê a nova interpretação, se houver.

12.3 Condições de segunda ordem

A introdução de um multiplicador de Lagrange como uma variável adicional possibilita a aplicação da mesma condição de primeira ordem usada no problema do extremo livre ao problema do extremo restrito. É tentador avançar mais um passo e também tomar emprestadas as condições de segunda ordem necessárias e suficientes. Contudo, isso não deve ser feito. Pois, ainda que Z^* seja, de fato, um tipo padrão de extremo no que diz respeito às variáveis de escolha, isso *não* acontece quando se trata do multiplicador de Lagrange. Especificamente, podemos ver, por (12.15) que, diferentemente de x^* e y^* , se λ^* for substituído por qualquer outro valor de λ , não produzirá nenhum efeito em Z^* , já que $[c - g(x^*, y^*)]$ é identicamente igual a zero. Assim, o papel desempenhado por λ na solução ótima difere basicamente do papel de x e y .[†] Conquanto não cause mal algum tratar λ como apenas mais uma variável de escolha na discussão da condição de primeira ordem, é preciso tomar cuidado e não aplicar cegamente as condições de segunda ordem desenvolvidas para o problema do extremo livre ao presente caso do extremo restrito. Em vez disso, devemos derivar um conjunto de novas condições. Como veremos, as novas condições podem ser enunciadas novamente em termos da diferencial total de segunda ordem d^2z . Contudo, a presença da restrição acarretará certas modificações significativas no critério.

[†] Em uma estrutura mais geral de otimização restrita, conhecida como “programação não-linear”, a ser discutida no Capítulo 13, mostraremos que, com restrições de desigualdade, se Z^* for um máximo (mínimo) em relação a x e y , então ele será, de fato, um mínimo (máximo) em relação a λ . Em outras palavras, o ponto (λ^*, x^*, y^*) é um ponto de sela. O presente caso – em que Z^* é um extremo genuíno em relação a x e y , mas é invariante em relação a λ – pode ser considerado um caso degenerado de ponto de sela. A natureza de ponto de sela da solução (λ^*, x^*, y^*) também leva ao importante conceito de “dualidade”. Mas esse assunto será tratado mais a fundo depois.

Diferencial total de segunda ordem

Já mencionamos que, considerando que a restrição $g(x, y) = c$ significa $dg = g_x dx + g_y dy = 0$, como em (12.10), dx e dy já não são mais ambas arbitrárias. É claro que ainda podemos considerar (digamos) dx como uma variação arbitrária, mas, então, dy deve ser considerada como dependente de dx e ser sempre escolhida de modo a satisfazer (12.10), isto é, satisfazer $dy = -(g_x/g_y) dx$. Visto de maneira diferente, uma vez especificado o valor de dx , dy dependerá de g_x e g_y , mas, já que as últimas derivadas dependem, por sua vez, das variáveis x e y , dy também dependerá de x e y . Então, é óbvio que a fórmula anterior para d^2z em (11.6), que é baseada na arbitrariedade de ambas, dx e dy , já não se aplica mais.

Para achar uma nova expressão adequada para d^2z , devemos tratar dy como uma variável dependente de x e y durante a diferenciação (se quisermos considerar dx uma constante). Assim,

$$\begin{aligned} d^2z &= d(dz) = \frac{\partial(dz)}{\partial x} dx + \frac{\partial(dz)}{\partial y} dy \\ &= \frac{\partial}{\partial x} (f_x dx + f_y dy) dx + \frac{\partial}{\partial y} (f_x dx + f_y dy) dy \\ &= \left[f_{xx} dx + \left(f_{xy} dy + f_y \frac{\partial dy}{\partial x} \right) \right] dx + \left[f_{yx} dx + \left(f_{yy} dy + f_y \frac{\partial dy}{\partial y} \right) \right] dy \\ &= f_{xx} dx^2 + f_{xy} dy dx + f_y \frac{\partial(dy)}{\partial x} dx + f_{yx} dx dy + f_{yy} dy^2 + f_y \frac{\partial(dy)}{\partial y} dy \end{aligned}$$

Uma vez que o terceiro e o sexto termos podem ser reduzidos a

$$f_y \left[\frac{\partial(dy)}{\partial x} dx + \frac{\partial(dy)}{\partial y} dy \right] = f_y d(dy) = f_y d^2y$$

a expressão desejada para d^2z é

$$d^2z = f_{xx} dx^2 + 2f_{xy} dx dy + f_{yy} dy^2 + f_y d^2y \quad (12.17)$$

que difere de (11.6) somente pelo último termo, $f_y d^2y$.

É preciso observar que este último termo é de *primeiro* grau [d^2y não é o mesmo que $(dy)^2$]; assim, sua presença em (12.17) desqualifica d^2z como uma forma quadrática. Todavia, d^2z pode ser transformada em uma forma quadrática em virtude da restrição $g(x, y) = c$. Posto que a restrição implica $dg = 0$ e também $d^2g = d(dg) = 0$, então, pelo procedimento usado para obter (12.17), podemos obter

$$(d^2g) = g_{xx} dx^2 + 2g_{xy} dx dy + g_{yy} dy^2 + g_y d^2y = 0$$

Resolvendo esta última equação para d^2y e substituindo o resultado em (12.17), podemos eliminar a expressão de primeiro grau d^2y e escrever d^2z como a seguinte forma quadrática:

$$d^2z = \left(f_{xx} - \frac{f_y}{g_y} g_{xx} \right) dx^2 + 2 \left(f_{xy} - \frac{f_y}{g_y} g_{xy} \right) dx dy + \left(f_{yy} - \frac{f_y}{g_y} g_{yy} \right) dy^2$$

Por causa de (12.11'), o primeiro coeficiente entre parenteses é redutível a $(f_{xx} - \lambda g_{xx})$, o mesmo se aplicando aos outros termos. Entretanto, diferenciando parcialmente as derivadas em (12.8), você constatará que as seguintes derivadas segundas

$$\begin{aligned}
 Z_{xx} &= f_{xx} - \lambda g_{xx} \\
 Z_{xy} &= f_{xy} - \lambda g_{xy} = Z_{yx} \\
 Z_{yy} &= f_{yy} - \lambda g_{yy}
 \end{aligned}
 \tag{12.18}$$

são exatamente iguais aos coeficientes entre parênteses. Portanto, fazendo uso da função de Lagrange, podemos finalmente expressar d^2z de modo mais elegante, ou seja,

$$\begin{aligned}
 d^2z &= Z_{xx} dx^2 + Z_{xy} dx dy \\
 &\quad + Z_{yx} dy dx + Z_{yy} dy^2
 \end{aligned}
 \tag{12.17'}$$

Os coeficientes de (12.17') são simplesmente as derivadas parciais segundas de Z em relação às variáveis de escolha x e y ; por conseguinte, juntas, elas podem dar origem a um determinante hessiano.

Condições de segunda ordem

Para um extremo restrito de $z = f(x, y)$, sujeito a $g(x, y) = c$, as condições necessárias e suficientes de segunda ordem ainda giram em torno do sinal algébrico da diferencial total de segunda ordem d^2z , calculada em um ponto estacionário. Contudo, há uma mudança importante. No presente contexto, estamos preocupados com a condição de sinal definido ou semidefinido de d^2z , *não* para todos os valores possíveis de dx e dy (não ambos igual a zero), mas *somente* para os valores de dx e dy (não ambos iguais a zero) que satisfaçam a restrição linear (12.10), $g_x dx + g_y dy = 0$. Assim, as condições *necessárias* de segunda ordem são:

Para o máximo de z : d^2z negativa semidefinida, sujeita a $dg = 0$

Para o mínimo de z : d^2z positiva semidefinida, sujeita a $dg = 0$

e as condições *suficientes* de segunda ordem são

Para o máximo de z : d^2z negativa definida, sujeita a $dg = 0$

Para o mínimo de z : d^2z positiva definida, sujeita a $dg = 0$

A seguir, vamos nos concentrar nas condições suficientes de segunda ordem.

Considerando que os pares (dx, dy) que satisfazem a restrição $g_x dx + g_y dy = 0$ constituem um mero subconjunto do conjunto de todos os dx e dy possíveis, a condição de sinal definido restrito é menos rigorosa – isto é, mais fácil de satisfazer – do que a condição de sinal definido restrito discutida no Capítulo 11. Em outras palavras, a condição suficiente de segunda ordem para um problema de extremo restrito é uma condição mais fraca que a condição para um problema de extremo livre. São boas notícias porque, diferentemente das condições necessárias, que devem ser rigorosas para que sirvam de mecanismos efetivos de filtragem, as condições suficientes devem ser fracas para que sejam verdadeiramente aproveitáveis.[†]

O hessiano aumentado

Como no caso do extremo livre, é possível expressar a condição suficiente de segunda ordem na forma determinante. Entretanto, no caso do extremo restrito, em lugar do determinante hessiano $|H|$ vamos encontrar o *hessiano aumentado*.

Como preparação para o desenvolvimento dessa idéia, vamos primeiramente analisar as condições para sinal definido de uma forma quadrática de duas variáveis, sujeita a uma restrição linear, digamos,

[†] “Uma conta bancária de um milhão de dólares” é claramente uma condição suficiente para “poder comer filé no jantar”. Mas a aplicabilidade extremamente limitada dessa condição a torna praticamente inútil. Uma condição suficiente mais significativa seria algo como “ter cinquenta dólares na carteira”, que é um requisito financeiro muito menos rigoroso.

$$q = au^2 + 2huv + bv^2 \quad \text{sujeita a} \quad \alpha u + \beta v = 0$$

Visto que a restrição implica $v = -(\alpha/\beta)u$, podemos reescrever q como uma função de apenas uma variável:

$$q = au^2 - 2h\frac{\alpha}{\beta}u^2 + b\frac{\alpha^2}{\beta^2} = (a\beta^2 - 2h\alpha\beta + b\alpha^2)\frac{u^2}{\beta^2}$$

É óbvio que q é positiva (negativa) definida se, e somente se, a expressão entre parênteses for positiva (negativa). Mas agora acontece que o seguinte determinante simétrico

$$\begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} = 2h\alpha\beta - a\beta^2 - b\alpha^2$$

é exatamente a *negativa* da citada expressão entre parênteses. Por consequência, podemos afirmar que

$$q \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ sujeita a } \alpha u + \beta v = 0 \text{ se, e somente se, } \begin{vmatrix} 0 & \alpha & \beta \\ \alpha & a & h \\ \beta & h & b \end{vmatrix} \begin{cases} < 0 \\ > 0 \end{cases}$$

Vale notar que o determinante usado nesse critério nada mais é que o discriminante da forma quadrática original $\begin{vmatrix} a & h \\ h & b \end{vmatrix}$, com uma linha adicionada na parte superior e uma coluna semelhante à esquerda. Além do mais, a adição é simplesmente composta dos dois coeficientes α e β da restrição, mais um zero na diagonal principal. Esse discriminante aumentado é simétrico.

EXEMPLO 1 Determine se $q = 4u^2 + 4uv + 3v^2$, sujeita a $u - 2v = 0$, é positiva definida ou negativa definida. Em

primeiro lugar, vamos formar o discriminante aumentado $\begin{vmatrix} 0 & 1 & -2 \\ 1 & 4 & 2 \\ -2 & 2 & 3 \end{vmatrix}$, que transformamos em simétrico dividindo o coeficiente de uv em duas partes iguais para inserir no determinante. Considerando que o determinante tem um valor negativo (-27), q é positiva definida.

Quando aplicadas à forma quadrática d^2z em (12.17'), as variáveis u e v tornam-se dx e dy , respectivamente, e o discriminante (simples) consiste no hessiano $\begin{vmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{vmatrix}$. Além disso, posto que a restrição à forma quadrática é $g_x dx + g_y dy = 0$, temos $\alpha = g_x$ e $\beta = g_y$. Assim, para valores de dx e dy que satisfazem a citada restrição, agora podemos ter o seguinte critério com determinante para a condição de sinal definido de d^2z :

$$d^2z \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ sujeita a } dg = 0 \text{ se, e somente se, } \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} \begin{cases} < 0 \\ > 0 \end{cases}$$

O determinante à direita, normalmente denominado um hessiano aumentado, será denotado por $|\bar{H}|$, em que a barra na parte superior simboliza o aumento. Tendo isso como base, podemos concluir que, dado um valor estacionário de $z = f(x, y)$ ou de $Z = f(x, y) + \lambda[c - g(x, y)]$, um $|\bar{H}|$ positivo é suficiente para estabelecê-lo como um máximo relativo de z ; de modo semelhante, um

$|\bar{H}|$ negativo é suficiente para estabelecê-lo como um mínimo – sendo todas as derivadas calculadas em $|\bar{H}|$ calculadas nos valores críticos de x e y .

Agora que já derivamos a condição suficiente de segunda ordem, é fácil verificar que, como afirmamos antes, a satisfação dessa condição garantirá que o jacobiano de variáveis endógenas (12.12) não se anule no estado ótimo. Substituindo (12.18) em (12.12) e multiplicando a primeira coluna e a primeira linha do jacobiano por -1 (o que não altera o valor do determinante), constatamos que

$$|J| = \begin{vmatrix} 0 & g_x & g_y \\ g_x & Z_{xx} & Z_{xy} \\ g_y & Z_{yx} & Z_{yy} \end{vmatrix} = |\bar{H}| \quad (12.19)$$

Isto é, o jacobiano de variáveis endógenas é idêntico ao hessiano aumentado – um resultado semelhante a (11.42), onde foi demonstrado que, no contexto do extremo livre, o jacobiano de variáveis endógenas é idêntico ao hessiano simples. Se, ao cumprir a condição suficiente, tivermos $|\bar{H}| \neq 0$ no ótimo, então $|J|$ também deve ser diferente de zero. Por consequência, ao aplicar o teorema da função implícita ao presente contexto, não seria totalmente fora de propósito substituir a condição $|\bar{H}| \neq 0$ pela condição usual $|J| \neq 0$. Essa prática será adotada quando analisarmos a estática comparativa de problemas de otimização restrita na Seção 12.5.

EXEMPLO 2

Agora vamos voltar ao Exemplo 1 da Seção 12.2 e averiguar se o valor estacionário encontrado ali resulta em um máximo ou um mínimo. Visto que $Z_x = y - \lambda$ e $Z_y = x - \lambda$, as derivadas parciais de segunda ordem são $Z_{xx} = 0$, $Z_{xy} = Z_{yx} = 1$, e $Z_{yy} = 0$. Os elementos novos de que precisamos são $g_x = 1$ e $g_y = 1$. Assim, constatamos que

$$|\bar{H}| = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2 > 0$$

o que estabelece o valor $z^* = 9$ como um máximo.

EXEMPLO 3

Continuando com o Exemplo 2 da Seção 12.2, vemos que $Z_1 = 2x_1 - \lambda$ e $Z_2 = 2x_2 - 4\lambda$. Esses resultam em $Z_{11} = 2$, $Z_{12} = Z_{21} = 0$ e $Z_{22} = 2$. Considerando a restrição $x_1 + 4x_2 = 2$, obtemos $g_1 = 1$ e $g_2 = 4$. Segue-se que o hessiano aumentado é

$$|\bar{H}| = \begin{vmatrix} 0 & 1 & 4 \\ 1 & 2 & 0 \\ 4 & 0 & 2 \end{vmatrix} = -34 < 0$$

e o valor de $z^* = \frac{4}{17}$ é um mínimo.

EXEMPLO 4

Considere um modelo simples de dois períodos no qual a utilidade de um consumidor é uma função do consumo em ambos os períodos. Seja a função utilidade do consumidor

$$U(x_1, x_2) = x_1 x_2$$

onde x_1 é o consumo no período 1 e x_2 é o consumo no período 2. O consumidor também dispõe de um orçamento B no início do período 1.

Seja r uma taxa de juro de mercado segundo a qual o consumidor pode tomar emprestado ou emprestar durante os dois períodos. A restrição orçamental intertemporal do consumidor é que x_1 e o valor presente de x_2 somem B . Assim,

$$x_1 + \frac{x_2}{1+r} = B$$

A função de Lagrange para esse problema de maximização de utilidade é

$$Z = x_1 x_2 + \lambda \left(B - x_1 - \frac{x_2}{1+r} \right)$$

com as condições de primeira ordem

$$\frac{\partial Z}{\partial \lambda} = B - x_1 - \frac{x_2}{1+r} = 0$$

$$\frac{\partial Z}{\partial x_1} = x_2 - \lambda = 0$$

$$\frac{\partial Z}{\partial x_2} = x_1 - \frac{\lambda}{1+r} = 0$$

Combinando as duas últimas equações de primeira ordem para eliminar λ , temos

$$\frac{x_2}{x_1} = \frac{\lambda}{\lambda / (1+r)} = 1+r$$

Substituindo essa equação na restrição orçamental, obtemos a solução

$$x_1^* = \frac{B}{2} \quad \text{e} \quad x_2^* = \frac{B(1+r)}{2}$$

Em seguida, devemos verificar a condição suficiente de segunda ordem para um máximo. O hessiano aumentado para esse problema é

$$|\bar{H}| = \begin{vmatrix} 0 & -1 & -\frac{1}{1+r} \\ -1 & 0 & 1 \\ -\frac{1}{1+r} & 1 & 0 \end{vmatrix} = \frac{2}{1+r} > 0$$

Assim, a condição suficiente de segunda ordem é satisfeita para um máximo U .

Caso de n Variáveis

Quando a função objetivo toma a forma

$$z = f(x_1, x_2, \dots, x_n) \quad \text{sujeita a} \quad g(x_1, x_2, \dots, x_n) = c$$

a condição de segunda ordem ainda depende do sinal de d^2z . Visto que esta última é uma forma quadrática restrita nas variáveis $dx_1, dx_2, \dots, dx_n$ sujeita à relação

$$(dg=) g_1 dx_1 + g_2 dx_2 + \dots + g_n dx_n = 0$$

as condições para positiva definida ou negativa definida para d^2z ser novamente envolvem um hessiano aumentado. Mas, dessa vez, essas condições devem ser expressas em termos dos menores principais líderes aumentados do hessiano.

Dado um hessiano aumentado

$$|\bar{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

seus menores principais líderes aumentados podem ser definidos como

$$|\bar{H}_2| \equiv \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & Z_{11} & Z_{12} \\ g_2 & Z_{21} & Z_{22} \end{vmatrix} \quad |\bar{H}_3| \equiv \begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & Z_{11} & Z_{12} & Z_{13} \\ g_2 & Z_{21} & Z_{22} & Z_{23} \\ g_3 & Z_{31} & Z_{32} & Z_{33} \end{vmatrix} \quad (\text{etc.})$$

sendo que o último é $|\bar{H}_n| = |\bar{H}|$. Nesses símbolos que acabamos de introduzir, a barra horizontal acima de H novamente significa aumentado, e o índice indica a ordem do menor principal líder que está sendo aumentado. Por exemplo, $|\bar{H}_2|$ envolve o segundo menor principal líder do hessiano (simples), aumentado com 0, g_1 e g_2 ; e o mesmo vale para os outros. As condições para positiva definida ou negativa definida de d^2z então são

$$d^2z \text{ é } \begin{cases} \text{positiva definida} \\ \text{negativa definida} \end{cases} \text{ sujeita a } dg = 0 \text{ se, e somente se, } \begin{cases} |\bar{H}_2|, |\bar{H}_3|, \dots, |\bar{H}_n| < 0 \\ |\bar{H}_2| > 0; |\bar{H}_3| < 0; |\bar{H}_4| > 0 \text{ etc.} \end{cases}$$

Na primeira, todos os menores principais líderes aumentados, começando com $|\bar{H}_2|$, devem ser negativos; na última, os sinais devem se alternar. Como antes, uma d^2z positiva definida é suficiente para estabelecer um valor estacionário de z como seu mínimo, ao passo que uma d^2z negativa definida é suficiente para estabelecer aquele valor como um máximo.

Juntando as pontas soltas da discussão, podemos resumir as condições para um extremo relativo restrito na Tabela 12.1. Temos de reconhecer, entretanto, que o critério enunciado na tabela não é completo. Como a condição suficiente de segunda ordem *não* é necessária, não satisfazer os critérios enunciados não exclui a possibilidade de que o valor estacionário seja, ainda assim, um máximo ou um mínimo. Entretanto, em muitas aplicações econômicas, essa condição suficiente de segunda ordem (relativamente menos rigorosa) ou é satisfeita ou supõe-se que seja satisfeita, de modo que as informações da tabela são adequadas. Achamos que será instrutivo você comparar os resultados contidos na Tabela 12.1 com os da Tabela 11.2 para o caso do extremo livre.

TABELA 12.1

Teste com determinantes para extremo relativo restrito: $z = f(x_1, x_2, \dots, x_n)$, sujeito a $g(x_1, x_2, \dots, x_n) = c$; com $Z = f(x_1, x_2, \dots, x_n) + \lambda[c - g(x_1, x_2, \dots, x_n)]$

Condição	Máximo	Mínimo
Condição necessária de primeira ordem	$Z_\lambda = Z_1 = Z_2 = \cdots = Z_n = 0$	$Z_\lambda = Z_1 = Z_2 = \cdots = Z_n = 0$
Condição suficiente de segunda ordem [†]	$ \bar{H}_2 > 0; \bar{H}_3 < 0; \dots; (-1)^n \bar{H}_n > 0$	$ \bar{H}_2 , \bar{H}_3 , \dots, \bar{H}_n < 0$

[†] Aplicável somente após a condição de primeira ordem ter sido satisfeita.

Caso de múltiplas restrições

Quando aparece mais de uma restrição no problema, a condição de segunda ordem envolve um hessiano aumentado por mais de uma linha e uma coluna. Suponha que há n variáveis de escolha e m restrições ($m < n$) da forma $g^j(x_1, \dots, x_n) = c_j$. Então, a função de Lagrange será

$$Z = f(x_1, \dots, x_n) + \sum_{j=1}^m \lambda_j [c_j - g^j(x_1, \dots, x_n)]$$

e o hessiano aumentado aparecerá como

$$|\bar{H}| \equiv \begin{vmatrix} 0 & 0 & \dots & 0 & g_1^1 & g_2^1 & \dots & g_n^1 \\ 0 & 0 & \dots & 0 & g_1^2 & g_2^2 & \dots & g_n^2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & g_1^m & g_2^m & \dots & g_n^m \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ g_1^1 & g_1^2 & \dots & g_1^m & Z_{11} & Z_{12} & \dots & Z_{1n} \\ g_2^1 & g_2^2 & \dots & g_2^m & Z_{21} & Z_{22} & \dots & Z_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ g_n^1 & g_n^2 & \dots & g_n^m & Z_{n1} & Z_{n2} & \dots & Z_{nn} \end{vmatrix}$$

onde $g_i^j \equiv \partial g^j / \partial x_i$ são as derivadas parciais das funções de restrição, e os símbolos Z com dois índices denotam, como antes, as derivadas parciais de segunda ordem da função de Lagrange. Note que subdividimos o hessiano aumentado em quatro *blocos* para uma visualização mais clara. O bloco superior esquerdo é composto somente de zeros, e o bloco inferior direito é apenas o hessiano simples. Os outros dois blocos, contendo as derivadas g_i^j , guardam uma relação de imagem especular entre si com referência à diagonal principal, resultando, assim, em um arranjo simétrico de elementos em todo o hessiano aumentado.

Vários menores principais líderes aumentados podem ser formados a partir de $|\bar{H}|$. O que contiver Z_{22} como o último elemento de sua diagonal principal pode ser denotado por $|\bar{H}_2|$, como antes. Incluindo mais uma linha e mais uma coluna, de modo que Z_{33} entre em cena, teremos $|\bar{H}_3|$ e assim por diante. Com essa simbologia, podemos enunciar a condição suficiente de segunda ordem em termos dos sinais dos seguintes $(n - m)$ menores principais líderes aumentados:

$$|\bar{H}_{m+1}|, |\bar{H}_{m+2}|, \dots, |\bar{H}_n| (= |\bar{H}|)$$

Para um máximo de z , uma condição suficiente é que esses menores principais líderes aumentados alternem sinais, sendo que o sinal de $|\bar{H}_{m+1}|$ é o de $(-1)^{m+1}$. Para um mínimo de z , uma condição suficiente é que todos esses menores principais tenham o mesmo sinal, a saber, o de $(-1)^m$.

Note que ter um número par ou um número ímpar de restrições faz uma grande diferença, porque (-1) elevado a uma potência ímpar resultará no sinal oposto ao de (-1) elevado a uma potência par. Note, também, que, quando $m = 1$, a condição que acabamos de enunciar se reduz à apresentada na Tabela 12.1.

EXERCÍCIO 12.3

1. Use o hessiano aumentado para determinar se o valor estacionário de z obtido em cada parte do Exercício 12.2-1 é um máximo ou um mínimo.
2. Quando enunciamos as condições suficientes de segunda ordem para máximo e mínimo restritos, especificamos os sinais algébricos de $|\bar{H}_2|$, $|\bar{H}_3|$, $|\bar{H}_4|$ etc., mas não o de $|\bar{H}_1|$. Escreva uma expressão adequada para $|\bar{H}_1|$ e verifique se ele assume invariavelmente o sinal negativo.

3. Recordando a Propriedade II de determinantes (Seção 5.3), mostre que:

(a) Trocando adequadamente duas linhas e/ou duas colunas de $|\bar{H}_2|$ e alterado devidamente o sinal do determinante após cada troca, ele pode ser transformado em

$$\begin{vmatrix} Z_{11} & Z_{12} & g_1 \\ Z_{21} & Z_{22} & g_2 \\ g_1 & g_2 & 0 \end{vmatrix}$$

(b) Por um procedimento semelhante, $|\bar{H}_3|$ pode ser transformado em

$$\begin{vmatrix} Z_{11} & Z_{12} & Z_{13} & g_1 \\ Z_{21} & Z_{22} & Z_{23} & g_2 \\ Z_{31} & Z_{32} & Z_{33} & g_3 \\ g_1 & g_2 & g_3 & 0 \end{vmatrix}$$

Qual é o modo alternativo de “aumentar” os menores principais do hessiano sugerido por esses resultados?

4. Escreva o hessiano aumentado para um problema de otimização restrita com quatro variáveis de escolha e duas restrições. Depois, enuncie especificamente a condição suficiente de segunda ordem para um máximo e para um mínimo de z , respectivamente.

12.4 Quase-concavidade e quase-convexidade

Na Seção 11.5, demonstramos que, para um problema de extremo livre, saber se uma função objetivo é côncava ou convexa elimina a necessidade de verificar a condição de segunda ordem. No contexto da otimização restrita, novamente é possível prescindir da condição de segunda ordem se a superfície ou a hipersuperfície tiver o tipo adequado de configuração. Mas, dessa vez, a configuração desejada é a quase-concavidade (em vez de concavidade) para um máximo, e a quase-convexidade (em vez de convexidade) para um mínimo. Como demonstraremos, a quase-concavidade (quase-convexidade) é uma condição mais fraca que a concavidade (convexidade). E isso era de se esperar, posto que a condição suficiente de segunda ordem da qual prescindiremos também é mais fraca para o problema de otimização restrita (d^2z definida em sinal apenas para os dx_i que satisfazem $dg=0$) do que para o de otimização livre (d^2z definida em sinal para *todos* dx_i).

Caracterização geométrica

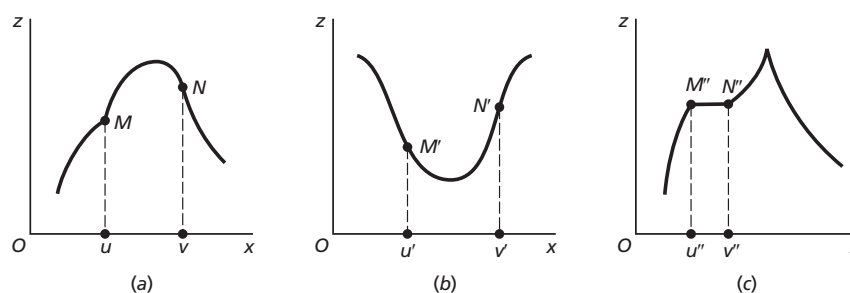
A quase-concavidade e a quase-convexidade, assim como a concavidade e a convexidade, podem ser ou estritas ou não-estritas. Em primeiro lugar, apresentaremos a caracterização geométrica desses conceitos:

Sejam u e v quaisquer dois pontos distintos no domínio (um conjunto convexo) de uma função f e considere o segmento de reta uv no domínio que dá origem ao arco MN no gráfico da função, tal que o ponto N é mais alto ou tenha a mesma altura do ponto M . Então, diz-se que f é *quase-côncava* (*quase-convexa*) se todos os pontos sobre o arco MN , exceto M e N , forem mais altos que ou tiverem a mesma altura do ponto M (mais baixos que ou tiverem a mesma altura do ponto N). Diz-se que a função f é *estritamente quase-côncava* (*estritamente quase-convexa*) se todos os pontos sobre o arco MN , exceto M e N , forem estritamente mais altos que o ponto M (estritamente mais baixos que o ponto N).

Isso deve deixar claro que qualquer função estritamente quase-côncava (estritamente quase-convexa) é quase-côncava (quase-convexa), mas a recíproca não é verdadeira.

Para ter uma noção melhor, vamos examinar as ilustrações da Figura 12.3, todas desenhadas para o caso de uma única variável. Na Figura 12.3a, o segmento de reta uv no domínio dá origem ao arco MN sobre a curva tal que N é mais alto que M . Visto que todos os pontos entre M e N sobre o

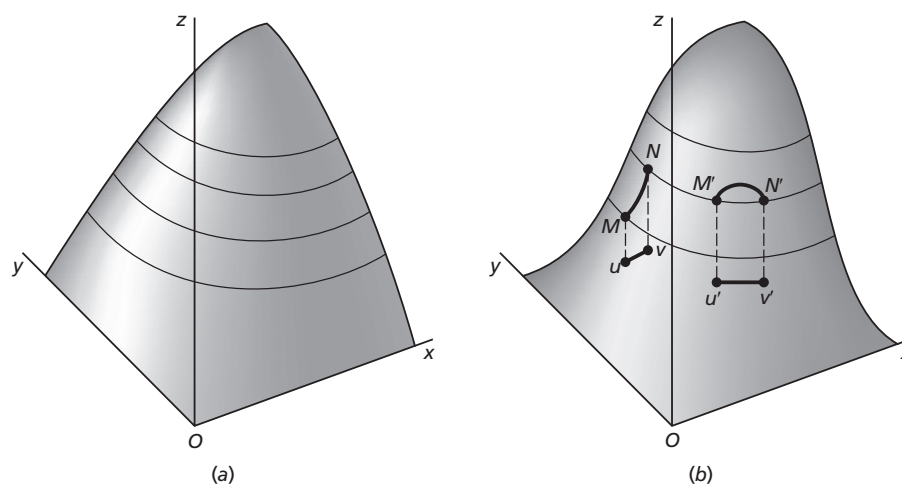
FIGURA 12.3



arco citado são estritamente mais altos que M , esse arco específico satisfaz a condição de quase-concavidade estrita. Para que a curva se qualifique como estritamente quase-côncava, contudo, *todos* os pares (u, v) possíveis devem ter arcos que satisfaçam a mesma condição. Esse é, de fato, o caso da função na Figura 12.3a. Note que essa função também satisfaz a condição de quase-concavidade (não-estrita). Porém, ela não cumpre a condição de quase-convexidade, porque alguns pontos sobre o arco MN são mais altos que N , o que é proibido para uma função quase-convexa. A função na Figura 12.3b tem a configuração oposta. Todos os pontos sobre o arco $M'N'$ são mais baixos que N' , a mais alta das duas extremidades, e o mesmo é verdadeiro para todos os arcos que podem ser desenhados. Assim, a função na Figura 12.3b é estritamente quase-convexa. Como você pode verificar, ela também satisfaz a condição de quase-convexidade (não-estrita), mas deixa de cumprir a condição de quase-concavidade. O que distingue a Figura 12.3c é a presença de um segmento de reta horizontal $M''N''$, no qual todos os pontos têm a mesma altura. O resultado é que aquele segmento de reta – e, por conseguinte, a curva inteira – só pode cumprir a condição de quase-concavidade, mas não a de quase-concavidade estrita.

Em termos gerais, o formato do gráfico de uma função quase-côncava que não é também côncava é aproximadamente um sino ou uma parte de um sino, e o formato do gráfico de uma função quase-convexa parece um sino invertido ou uma porção de um sino invertido. É admissível (mas não requerido) ter segmentos côncavos e convexos no sino. Essa natureza mais permissiva da caracterização faz da quase-concavidade (quase-convexidade) uma condição mais fraca que a concavidade (convexidade). Na Figura 12.4, comparamos a concavidade estrita com a quase-concavidade estrita para o caso de duas variáveis. Nesse desenho, ambas as superfícies retratam funções crescentes, pois contêm todas as porções ascendentes de um domo e de um sino, respectivamente. A superfície na Figura 12.4a é estritamente côncava, mas a da Figura 12.4b com certeza não é, pois contém porções convexas perto da base do sino. Ainda assim, ela é estritamente quase-côncava; todos os arcos sobre a superfície, exemplificados por MN e $M'N'$, satisfazem a condição de que todos os pontos sobre cada arco entre os dois pontos das extremidades são mais altos que o ponto da extremidade mais baixa. Voltando à Figura 12.4a, devemos notar que a superfície nela representada também é estritamente quase-côncava. Embora não tenhamos dese-

FIGURA 12.4



nhado nenhum arco ilustrativo MN e $M'N'$ na Figura 12.4a, não é difícil verificar que todos os arcos possíveis realmente satisfazem a condição de quase-concavidade estrita. Em geral, uma função estritamente côncava deve ser estritamente quase-côncava, embora a recíproca não seja verdadeira. Nos parágrafos seguintes, demonstraremos isso de maneira mais formal.

Definição algébrica

A caracterização geométrica precedente pode ser traduzida para uma definição algébrica de modo a facilitar a generalização para casos de um número de dimensões mais alto:

Uma função f é $\begin{cases} \text{quase-côncava} \\ \text{quase-convexa} \end{cases}$ se, e somente se, para qualquer par de pontos distintos u e v no domínio (conjunto convexo) de f , e para $0 < \theta < 1$,

$$f(v) \geq f(u) \Rightarrow f[\theta u + (1 - \theta)v] \begin{cases} \geq f(u) \\ \leq f(v) \end{cases} \quad (12.20)$$

Para adaptar essa definição à quase-concavidade e à quase-convexidade *estritas*, as duas desigualdades fracas da direita devem ser trocadas para desigualdades estritas $\begin{cases} > f(u) \\ < f(v) \end{cases}$. Será instrutivo comparar (12.20) com (11.20).

A partir dessa definição, seguem-se, de pronto, os três teoremas seguintes. Eles serão enunciados em termos de uma função $f(x)$, onde x pode ser interpretado como um vetor de variáveis, $x = (x_1, \dots, x_n)$.

Teorema I (negativa de uma função) Se $f(x)$ for quase-côncava (estritamente quase-côncava), então $-f(x)$ é quase-convexa (estritamente quase-convexa).

Teorema II (concavidade versus quase-concavidade) Qualquer função côncava (convexa) é quase-côncava (quase-convexa), mas a recíproca não é verdadeira. De modo semelhante, qualquer função estritamente côncava (estritamente convexa) é estritamente quase-côncava (estritamente quase-convexa), mas a recíproca não é verdadeira.

Teorema III (função linear) Se $f(x)$ for uma função linear, então ela é quase-côncava, bem como quase-convexa.

O teorema I resulta do fato de que multiplicar uma desigualdade por -1 inverte o sentido da desigualdade. Seja $f(x)$ quase-côncava, com $f(v) \geq f(u)$. Então, por (12.20), $f[\theta u + (1 - \theta)v] \geq f(u)$. No que diz respeito à função $-f(x)$, todavia, temos (após multiplicar as duas desigualdades por -1) $-f(u) \geq -f(v)$ e $-f[\theta u + (1 - \theta)v] \leq -f(u)$. Interpretando $-f(u)$ como a altura do ponto N , e $-f(v)$ como a altura de M , vemos que a função $-f(x)$ satisfaz a condição de quase-convexidade em (12.20). Isso prova um dos quatro casos citados no Teorema I; as demonstrações para os outros três são semelhantes.

Para o Teorema II, provaremos apenas que a concavidade implica a quase-concavidade. Seja $f(x)$ côncava. Então, por (11.20),

$$f[\theta u + (1 - \theta)v] \geq \theta f(u) + (1 - \theta)f(v)$$

Agora, suponha que $f(v) \geq f(u)$; então, qualquer média ponderada de $f(v)$ e $f(u)$ não tem nenhuma possibilidade de ser menor que $f(u)$, isto é,

$$\theta f(u) + (1 - \theta)f(v) \geq f(u)$$

Combinando esses dois resultados, constatamos, por transitividade, que

$$f[\theta u + (1 - \theta)v] \geq f(u) \quad \text{para } f(v) \geq f(u)$$

o que satisfaz a definição de quase-concavidade em (12.20). Note, todavia, que a condição de quase-concavidade não pode garantir a concavidade.

Uma vez estabelecido o Teorema II, o Teorema III segue-se imediatamente. Já sabemos que a função linear é côncava e convexa, embora não estritamente. Em vista do Teorema II, uma função linear também deve ser tanto quase-côncava como quase-convexa, embora não estritamente.

No caso de funções côncavas e convexas, há um teorema útil no sentido de que a soma de funções côncavas (convexas) também é côncava (convexa). Infelizmente, esse teorema não pode ser generalizado para funções quase-côncavas e quase-convexas. Por exemplo, uma soma de duas funções quase-côncavas *não é necessariamente* quase-côncava (veja Exercício 12.4-3).

Às vezes pode ser mais fácil verificar a quase-concavidade e a quase-convexidade pela seguinte definição alternativa:

Uma função $f(x)$, onde x é um vetor de variáveis, é $\begin{cases} \text{quase-côncava} \\ \text{quase-convexa} \end{cases}$ se, e somente se, para qualquer constante k , o conjunto

$$\begin{cases} S^{\geq} \equiv \{x | f(x) \geq k\} \\ S^{\leq} \equiv \{x | f(x) \leq k\} \end{cases} \text{ for um conjunto convexo.} \quad (12.21)$$

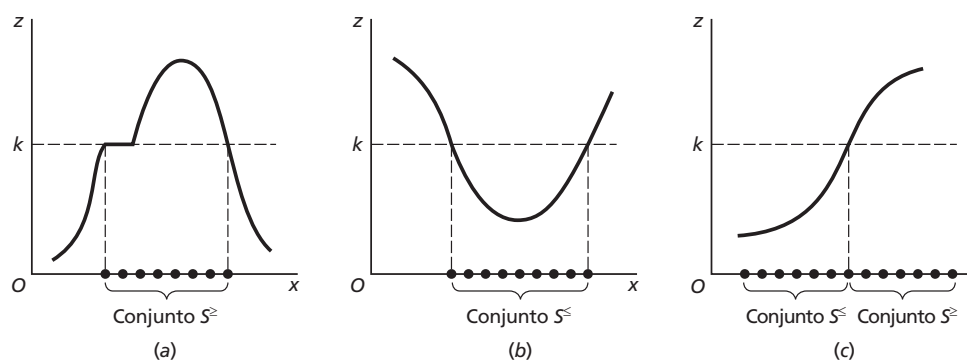
Os conjuntos $S^{\geq}$ e $S^{\leq}$, que são subconjuntos do domínio, foram introduzidos anteriormente (Figura 11.10) para mostrar que uma função convexa (ou mesmo uma função côncava) pode dar origem a um conjunto convexo. Aqui, estamos empregando esses dois conjuntos como testes de quase-concavidade e quase-convexidade. As três funções da Figura 12.5 contêm segmentos côncavos, bem como segmentos convexas e, portanto, não são convexas nem côncavas. Mas a função na Figura 12.5a é quase-côncava, porque, para qualquer valor de k (dos quais apenas um foi ilustrado), o conjunto $S^{\geq}$ é convexo. A função na Figura 12.5b é, por outro lado, quase-convexa, visto que o conjunto $S^{\leq}$ é convexo. A função na Figura 12.5c – uma função monotônica – difere das outras duas no sentido de que ambos, $S^{\geq}$ e $S^{\leq}$, são conjuntos convexas. Por conseguinte, aquela função é quase-côncava e quase-convexa.

Note que, conquanto (12.21) possa ser usada para verificar a quase-concavidade e a quase-convexidade, ela é incapaz de distinguir entre variedades estritas e não-estritas dessas propriedades. Note, também, que as propriedades definidoras em (12.21) não são, em si, suficientes para concavidade e convexidade, respectivamente. Em particular, dada uma função côncava que deve ser forçosamente quase-côncava, podemos concluir que $S^{\geq}$ é um conjunto convexo; mas, dado que $S^{\leq}$ é um conjunto convexo, podemos concluir apenas que a função f é quase-côncava (mas não necessariamente côncava).

EXEMPLO 1

Verifique a quase-concavidade e a quase-convexidade de $z = x^2$ ($x \geq 0$). É fácil verificar geometricamente que essa função é convexa; na verdade, estritamente convexa. Portanto, ela é quase-convexa. O interessante é que ela também é quase-côncava. Isso porque seu gráfico – a metade direita de uma curva em U, partindo do ponto de origem e aumentando a uma taxa crescente – é, de modo semelhante à Figura 12.5c, capaz de gerar um $S^{\geq}$ convexo, bem como um $S^{\leq}$ convexo.

FIGURA 12.5



Se, em vez disso, quisermos aplicar (12.20), deixaremos primeiramente que u e v sejam dois valores distintos não-negativos de x . Então

$$f(u) = u^2 \quad f(v) = v^2 \quad \text{e} \quad f[\theta u + (1 - \theta)v] = [\theta u + (1 - \theta)v]^2$$

Suponha que $f(v) \geq f(u)$, isto é, $v^2 \geq u^2$; então $v \geq u$ ou, mais especificamente, $v > u$ (já que u e v são distintas). Considerando que a média ponderada $[\theta u + (1 - \theta)v]$ deve estar entre u e v , podemos escrever a desigualdade contínua

$$v^2 > [\theta u + (1 - \theta)v]^2 > u^2 \quad \text{para } 0 < \theta < 1$$

ou

$$f(v) > f[\theta u + (1 - \theta)v] > f(u) \quad \text{para } 0 < \theta < 1$$

Por (12.20), esse resultado torna a função f quase-côncava e quase-convexa – na verdade, estritamente.

EXEMPLO 2

Mostre que $z = f(x, y) = xy$ (com $x, y \geq 0$) é quase-côncava. Usaremos o critério em (12.21) e estabeleceremos que o conjunto $S^z = \{(x, y) \mid xy \geq k\}$ é um conjunto convexo para qualquer k . Para esse propósito, estabelecemos $xy = k$ para obter uma curva de isovalor para cada valor de k . Assim como x e y , k deve ser não-negativo. No caso de $k > 0$, a curva de isovalor é uma hipérbole retangular no primeiro quadrante do plano xy . O conjunto S^z , composto de todos os pontos sobre ou acima de uma hipérbole retangular, é um conjunto convexo. No outro caso, com $k = 0$, a curva de isovalor, como definida por $xy = 0$, tem forma de L, sendo que o L coincide com os segmentos não-negativos dos eixos x e y . O conjunto S^z , consistindo, dessa vez, em todo o quadrante não-negativo é, mais uma vez, um conjunto convexo. Assim, por (12.21), a função $z = xy$ (com $x, y \geq 0$) é quase-côncava.

É preciso tomar cuidado para não confundir a forma das curvas de isovalor $xy = k$ (que são definidas no plano xy) com a forma da superfície $z = xy$ (que é definida no espaço xyz). A característica da superfície z (quase-côncava no espaço tridimensional) é o que queremos averiguar; a forma das curvas de isovalor (convexa no espaço bidimensional para k positivo) nos interessa aqui apenas como um meio de delinear os conjuntos S^z de modo a aplicar o critério em (12.21).

EXEMPLO 3

Mostre que $z = f(x, y) = (x - a)^2 + (y - b)^2$ é quase-convexa. Vamos aplicar novamente (12.21). Estabelecendo $(x - a)^2 + (y - b)^2 = k$, vemos que k deve ser não-negativo. Para cada k , a curva de isovalor é um círculo no plano xy cujo centro é (a, b) e com raio $\sqrt{k}$. Visto que $S^z = \{(x, y) \mid (x - a)^2 + (y - b)^2 \leq k\}$ é o conjunto de todos os pontos sobre ou dentro de um círculo, ele constitui um conjunto convexo. Isso é verdade mesmo quando $k = 0$ – quando o círculo degenera para um único ponto, (a, b) – já que, por convenção, um ponto único é considerado um conjunto convexo. Assim, a função dada é quase-convexa.

Funções diferenciáveis

As definições (12.20) e (12.21) não requerem diferenciabilidade da função f . Se f for diferenciável, contudo, a quase-concavidade e a quase-convexidade podem ser definidas alternativamente em termos de suas derivadas primeiras:

Uma função diferenciável de uma só variável, $f(x)$, é $\begin{cases} \text{quase-côncava} \\ \text{quase-convexa} \end{cases}$ se, e somente se, para qualquer par de pontos distintos u e v no domínio,

$$f(v) \geq f(u) \Rightarrow \begin{cases} f'(u)(v - u) \\ f'(v)(v - u) \end{cases} \geq 0 \quad (12.22)$$

A quase-concavidade e a quase-convexidade serão *estritas*, se a desigualdade fraca à direita for trocada para a desigualdade estrita > 0 . Quando há duas ou mais variáveis independentes, a definição deve ser modificada da seguinte maneira:

Uma função diferenciável $f(x_1, \dots, x_n)$ é $\begin{cases} \text{quase-concava} \\ \text{quase-convexa} \end{cases}$ se, e somente se, para quaisquer dois pontos distintos $u = (u_1, \dots, u_n)$ e $v = (v_1, \dots, v_n)$ no domínio,

$$f(v) \geq f(u) \Rightarrow \left\{ \begin{array}{l} \sum_{j=1}^n f_j(u)(v_j - u_j) \\ \sum_{j=1}^n f_j(v)(v_j - u_j) \end{array} \right\} \geq 0 \quad (12.22')$$

onde $f_j \equiv \partial f / \partial x_j$ deve ser calculada em u ou v conforme o caso.

Mais uma vez, para a quase-concavidade e a quase-convexidade *estritas*, a desigualdade fraca da direita deve ser mudada para a desigualdade estrita > 0 .

Finalmente, se uma função $z = f(x_1, \dots, x_n)$ for continuamente diferenciável duas vezes, a quase-concavidade e a quase-convexidade podem ser verificadas por meio das derivadas parciais primeiras e segundas da função, arranjadas no determinante aumentado

$$|B| = \begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ f_n & f_{n1} & f_{n2} & \cdots & f_{nn} \end{vmatrix} \quad (12.23)$$

Esse determinante aumentado é parecido com o hessiano aumentado $|\overline{H}|$ introduzido na Seção 12.3. Mas, diferentemente do último, o acréscimo em $|B|$ é composto das derivadas primeiras da função f em vez de uma função restrita estranha g . Devido ao fato de que $|B|$ depende exclusivamente das derivadas da função f em si, podemos usar $|B|$, juntamente com seus menores principais líderes

$$|B_1| = \begin{vmatrix} 0 & f_1 \\ f_1 & f_{11} \end{vmatrix} \quad |B_2| = \begin{vmatrix} 0 & f_1 & f_2 \\ f_1 & f_{11} & f_{12} \\ f_2 & f_{21} & f_{22} \end{vmatrix} \quad \dots \quad |B_n| = |B| \quad (12.24)$$

para caracterizar a configuração daquela função.

Enunciaremos aqui duas condições; uma é necessária e a outra é suficiente. Ambas estão relacionadas com a quase-concavidade em um domínio que consiste apenas no *ortante não negativo* (o análogo de n dimensões do quadrante não negativo), isto é, com $x_1, \dots, x_n \geq 0$.[†]

Para que $z = f(x_1, \dots, x_n)$ seja quase-concava no ortante não negativo, é necessário que

$$|B_1| \leq 0, \quad |B_2| \geq 0, \quad \dots, \quad |B_n| \begin{cases} \leq \\ \geq \end{cases} 0 \text{ se } n \text{ for } \begin{cases} \text{ímpar} \\ \text{par} \end{cases} \quad (12.25)$$

onde quer que as derivadas parciais sejam calculadas no ortante não negativo.

[†] Conquanto a concavidade (convexidade) de uma função em um domínio convexo sempre possa ser estendida para a concavidade (convexidade) em todo o espaço, isso não acontece com a quase-concavidade e a quase-convexidade. Por exemplo, nossas conclusões nos Exemplos 1 e 2 não serão válidas se for permitido que as variáveis assumam valores negativos. As duas condições dadas aqui são baseadas em Arrow, Kenneth J. e Enthoven, Alain C. "Quasi-Concave Programming". *Econometrica*, p. 797, out. 1961 (Teorema 5), e Takayama, Akira. *Analytical Methods in Economics*. University of Michigan Press, 1993, p. 65 (Teorema 1.12).

Uma condição suficiente para f ser estritamente quase-côncava no ortante não negativo é que

$$|B_1| < 0, \quad |B_2| > 0, \quad \dots, \quad |B_n| \begin{cases} < \\ > \end{cases} 0, \quad \text{se } n \text{ for } \begin{cases} \text{ímpar} \\ \text{par} \end{cases} \quad (12.26)$$

onde quer que as derivadas parciais sejam calculadas no ortante não negativo.

Note que a condição $|B_1| \leq 0$ em (12.25) é automaticamente satisfeita porque $|B_1| = -f_1^2$; ela está citada aqui apenas por questão de simetria, o mesmo acontecendo com a condição $|B_1| < 0$ em (12.26).

EXEMPLO 4

A função $z = f(x_1, x_2) = x_1 x_2$ ($x_1, x_2 \geq 0$) é quase-côncava (veja Exemplo 2). Agora vamos verificar essa característica por (12.22'). Sejam $u = (u_1, u_2)$ e $v = (v_1, v_2)$ dois pontos quaisquer no domínio. Então $f(u) = u_1 u_2$ e $f(v) = v_1 v_2$. Suponha que

$$f(v) \geq f(u) \quad \text{ou} \quad v_1 v_2 \geq u_1 u_2 \quad (v_1, v_2, u_1, u_2 \geq 0) \quad (12.27)$$

Visto que as derivadas parciais de f são $f_1 = x_2$ e $f_2 = x_1$, (12.22') equivale à condição

$$f_1(u)(v_1 - u_1) + f_2(u)(v_2 - u_2) = u_2(v_1 - u_1) + u_1(v_2 - u_2) \geq 0$$

ou, após rearranjo,

$$u_2(v_1 - u_1) \geq u_1(u_2 - v_2) \quad (12.28)$$

Precisamos considerar quatro possibilidades em relação aos valores de u_1 e u_2 . Primeiro, se $u_1 = u_2 = 0$, então (12.28) é trivialmente satisfeita. Segundo, se $u_1 = 0$, mas $u_2 > 0$, então (12.28) se reduz à condição $u_2 v_1 \geq 0$, que é novamente satisfeita, visto que u_2 e v_1 são ambas não negativas. Terceiro, se $u_1 > 0$ e $u_2 = 0$, então (12.28) se reduz à condição $0 \geq -u_1 v_2$, que ainda é satisfeita. Quarto e último, suponha que u_1 e u_2 são ambas positivas, de modo que v_1 e v_2 também são positivas. Subtraindo $v_2 u_1$ de ambos os lados de (12.27), obtemos

$$v_2(v_1 - u_1) \geq u_1(u_2 - v_2) \quad (12.29)$$

Agora apresentam-se três outras subpossibilidades:

1. Se $u_2 = v_2$, então $v_1 \geq u_1$. De fato, devemos ter $v_1 > u_1$ visto que (u_1, u_2) e (v_1, v_2) são pontos distintos. O fato de que $u_2 = v_2$ e $v_1 > u_1$ implica que a condição (12.28) é satisfeita.
2. Se $u_2 > v_2$, então também devemos ter $v_1 > u_1$ por (12.29). Multiplicando ambos os lados de (12.29) por u_2/v_2 , obtemos

$$u_2(v_1 - u_1) \geq \frac{u_2}{v_2} u_1(u_2 - v_2) > u_1(u_2 - v_2) \quad \left[\text{já que } \frac{u_2}{v_2} > 1 \right] \quad (12.30)$$

Assim, (12.28) é novamente satisfeita.

3. A subpossibilidade final é que $u_2 < v_2$, implicando que u_2/v_2 é uma fração positiva. Nesse caso, a primeira desigualdade de (12.30) ainda vale. A segunda desigualdade também vale, mas agora por uma razão diferente: uma fração (u_2/v_2) de um número negativo $(u_2 - v_2)$ é maior que esse número em si.

Considerando que (12.28) é satisfeita em toda situação possível, a função $z = x_1 x_2$ ($x_1, x_2 \geq 0$) é quase-côncava. Por conseguinte, a condição necessária (12.25) deve valer. Como as derivadas parciais de f são

$$f_1 = x_2 \quad f_2 = x_1 \quad f_{11} = f_{22} = 0 \quad f_{12} = f_{21} = 1$$

constatamos que os menores principais líderes relevantes são

$$|B_1| = \begin{vmatrix} 0 & x_2 \\ x_2 & 0 \end{vmatrix} = -x_2^2 \leq 0 \quad |B_2| = \begin{vmatrix} 0 & x_2 & x_1 \\ x_2 & 0 & 1 \\ x_1 & 1 & 0 \end{vmatrix} = 2x_1x_2 \geq 0$$

Assim, (12.25) é, de fato, satisfeita. Note, contudo, que a condição suficiente (12.26) é satisfeita somente no ortante positivo.

EXEMPLO 5

Mostre que $z = f(x, y) = x^a y^b$ ($x, y > 0$; $0 < a, b < 1$) é estritamente quase-côncava. As derivadas parciais dessa função são

$$\begin{aligned} f_x &= ax^{a-1}y^b & f_y &= bx^ay^{b-1} \\ f_{xx} &= a(a-1)x^{a-2}y^b & f_{xy} = f_{yx} &= abx^{a-1}y^{b-1} & f_{yy} &= b(b-1)x^ay^{b-2} \end{aligned}$$

Assim, os menores principais líderes de $|B|$ têm os seguintes sinais:

$$\begin{aligned} |B_1| &= \begin{vmatrix} 0 & f_x \\ f_x & f_{xx} \end{vmatrix} = -(ax^{a-1}y^b)^2 < 0 \\ |B_2| &= \begin{vmatrix} 0 & f_x & f_y \\ f_x & f_{xx} & f_{xy} \\ f_y & f_{yx} & f_{yy} \end{vmatrix} = [2a^2b^2 - a(a-1)b^2 - a^2b(b-1)]x^{3a-2}y^{3b-2} > 0 \end{aligned}$$

Isso satisfaz a condição suficiente para a quase-concavidade estrita em (12.26).

Um exame mais aprofundado do hessiano aumentado

O determinante aumentado $|B|$, como definido em (12.23), é diferente do hessiano aumentado

$$|\bar{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix}$$

de duas maneiras: (1) os elementos acrescidos em $|B|$ são as derivadas parciais de primeira ordem da função f , e não da função g ; e (2) os elementos restantes em $|B|$ são as derivadas parciais de segunda ordem de f em vez da função de Lagrange Z . Contudo, no caso especial de uma equação linear de restrição, $g(x_1, \dots, x_n) = a_1x_1 + \cdots + a_nx_n = c$ — um caso que se encontra com frequência na economia (veja Seção 12.5) — Z_{ij} se reduz a f_{ij} . Pois, então, a função de Lagrange é

$$Z = f(x_1, \dots, x_n) + \lambda(c - a_1x_1 - \cdots - a_nx_n)$$

de modo que

$$Z_j = f_j - \lambda a_j \quad \text{e} \quad Z_{ij} = f_{ij}$$

Examinando os acréscimos, notamos que a função restrita linear resulta na derivada primeira $g_j = a_j$. Além disso, quando a condição de primeira ordem é satisfeita, temos $Z_j = f_j - \lambda a_j = 0$, de modo que $f_j = \lambda a_j$ ou $f_j = \lambda g_j$. Assim, os acréscimos em $|B|$ são simplesmente os de $|\bar{H}|$ multiplicados por um escalar positivo λ . Fatorando λ sucessivamente nas linhas e colunas acrescidas de $|\bar{H}|$ (ver Seção 5.3, Exemplo 5), temos

$$|B| = \lambda^2 |\bar{H}|$$

Conseqüentemente, no caso da restrição linear, os dois determinantes aumentados sempre possuem o mesmo sinal no ponto estacionário de Z . Pelo mesmo critério, os menores principais líderes $|B_i|$ e $|\bar{H}_i|$ ($i = 1, \dots, n$) também devem compartilhar o mesmo sinal naquele ponto. Deduz-se, então, que, se o determinante aumentado $|B|$ satisfizer a condição suficiente para a quase-concavidade estrita em (12.26), o hessiano aumentado $|\bar{H}|$ deve, então, satisfazer a condição suficiente de segunda ordem para a maximização restrita na Tabela 12.1.

Extremos absolutos versus extremos relativos

Um quadro mais abrangente da relação entre a quase-concavidade e as condições de segunda ordem é apresentado na Figura 12.6. (Uma modificação adequada adaptará a figura para a quase-convexidade.) Construída segundo o mesmo critério da Figura 11.5 – e para ser lida da mesma maneira – essa figura relaciona a quase-concavidade com o máximo restrito *absoluto*, bem como *relativo*, de uma função diferenciável duas vezes $z = f(x_1, \dots, x_n)$. Os três acréscimos na parte superior resumem as condições de primeira e de segunda ordem para um máximo relativo restrito. E os retângulos da coluna do meio, assim como os da Figura 11.5, vinculam entre si os conceitos de máximo relativo, máximo absoluto e máximo absoluto único.

Mas as informações realmente interessantes estão nos dois losangos e nos símbolos $\Rightarrow$ alongados que passam por eles. O da esquerda nos informa que, uma vez satisfeita a condição de primeira ordem, e se as duas cláusulas relacionadas no losango também forem satisfeitas, temos uma condição suficiente para um máximo absoluto restrito. A primeira cláusula é que a função f seja explicitamente quase-côncava – um novo termo que definiremos imediatamente.

Uma função quase-côncava f é *explicitamente quase-côncava* se tiver a seguinte propriedade adicional

$$f(v) > f(u) \Rightarrow f[\theta u + (1 - \theta)v] > f(u)$$

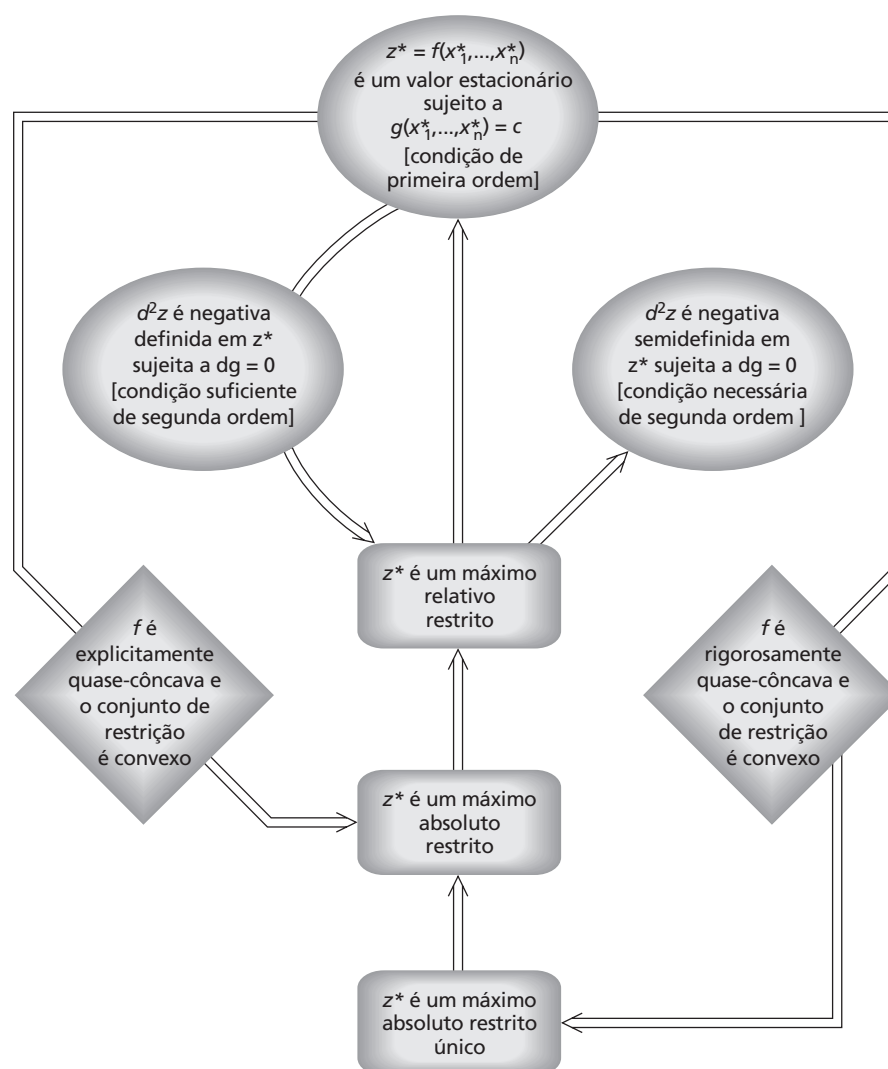
Essa propriedade definidora significa que, sempre que um ponto sobre a superfície, $f(v)$, for mais alto que um outro, $f(u)$, então todos os pontos intermediários – os pontos que estão sobre a superfície diretamente acima do segmento de reta uv no domínio – também devem ser mais altos que $f(u)$. Essa estipulação exclui quaisquer segmentos de planos *horizontais* sobre a superfície, exceto um platô no topo da superfície.[†] Note que a condição para a quase-concavidade *explícita* não é tão forte quanto a condição para a quase-concavidade *estrita*, visto que esta última requer que $f[\theta u + (1 - \theta)v] > f(u)$ mesmo para $f(v) = f(u)$, implicando que segmentos de plano *não horizontais* também são excluídos.[‡] A outra cláusula no losango do lado esquerdo é que o conjunto $\{(x_1, \dots, x_n) \mid g(x_1, \dots, x_n) = c\}$ seja convexo. Quando ambas as cláusulas são cumpridas, estaremos tratando com a porção de uma superfície (ou hipersuperfície) em formato de sino, livre de segmentos horizontais, que está diretamente acima de um conjunto convexo estabelecido no domínio. Um máximo local encontrado em tal subconjunto da superfície deve ser um máximo absoluto restrito.

O losango à direita na Figura 12.6 envolve a condição mais forte de quase-concavidade *estrita*. Uma função estrita quase-côncava deve ser explicitamente quase-côncava, embora a recíproca não seja verdadeiramente. Por conseguinte, quando a quase-concavidade estrita substitui a quase-concavidade explícita, ainda assim está assegurado um máximo absoluto restrito. Mas, dessa vez, esse máximo absoluto restrito também deve ser único, visto que a ausência de qualquer segmento de plano em qualquer lugar sobre a superfície exclui definitivamente a possibilidade de vários máximos restritos.

[†] Considere uma superfície que contém um segmento de plano horizontal P tal que $f(u) \in P$ e $f(v) \notin P$. Então, esses pontos intermediários que estão localizados sobre P terão a mesma altura de $f(u)$, violando, portanto, a primeira cláusula.

[‡] Considere uma superfície que contém um segmento de plano inclinado P' tal que $f(u) = f(v)$ estão ambos localizados sobre P' . Então, todos os pontos intermediários também estarão sobre P' e terão a mesma altura de $f(u)$, violando, portanto, o requisito citado de quase-concavidade estrita.

FIGURA 12.6

**EXERCÍCIO 12.4**

- Trace uma curva estritamente quase-côncava $z = f(x)$ que é
 - também quase-convexa
 - não quase-convexa
 - não convexa
 - não-côncava
 - nem côncava nem convexa
 - côncava e também convexa
- As seguintes funções são quase-côncavas? Estritamente? Em primeiro lugar, faça a verificação gráfica e, em seguida, a algébrica por (12.20). Suponha que $x \geq 0$.
 - $f(x) = a$
 - $f(x) = a + bx$ ($b > 0$)
 - $f(x) = a + cx^2$ ($c < 0$)
- Seja $z = f(x)$ uma função cujo gráfico é uma curva de inclinação negativa de formato semelhante à metade direita de um sino no primeiro quadrante, contendo os pontos (0, 5), (2, 4), (3, 2) e (5, 1). Seja $z = g(x)$ uma função cujo gráfico é uma reta de inclinação positiva a 45° . As funções $f(x)$ e $g(x)$ são quase-côncavas?
 - Agora represente em gráfico a soma $f(x) + g(x)$. A função soma é quase-côncava?
- Examinando seus gráficos e usando (12.21), verifique se as seguintes funções são quase-côncavas, quase-convexas, ambas ou nenhuma:
 - $f(x) = x^3 - 2x$
 - $f(x_1, x_2) = 6x_1 - 9x_2$
 - $f(x_1, x_2) = x_2 - \ln x_1$

5. (a) Verifique que a função cúbica $z = ax^3 + bx^2 + cx + d$ não é, em geral, quase-côncava nem quase-convexa.
 (b) É possível impor restrições sobre os parâmetros de modo que a função se torne quase-côncava e quase-convexa para $x \geq 0$?
6. Use (12.22) para verificar a quase-concavidade e a quase-convexidade de $z = x^2 (x \geq 0)$.
7. Mostre que $z = xy$ ($x, y \geq 0$) não é quase-convexa.
8. Use determinantes aumentados para verificar a quase-concavidade e a quase-convexidade das seguintes funções:
 (a) $z = -x^2 - y^2$ ($x, y > 0$) (b) $z = -(x+1)^2 - (y+2)^2$ ($x, y > 0$)

12.5 Maximização da utilidade e demanda do consumidor

A maximização de uma função utilidade foi citada na Seção 12.1 como um exemplo de otimização restrita. Agora vamos examinar novamente esse problema com maiores detalhes. Por simplicidade, ainda supomos que nosso consumidor pode escolher hipoteticamente entre somente dois bens e que ambos têm funções utilidade marginal contínuas e positivas. Os preços dos dois bens são determinados pelo mercado, portanto, são exógenos, mas nesta seção vamos omitir o índice zero dos símbolos de preço. Se o poder de compra do consumidor for uma dada quantidade B (de “budget” – orçamento), o problema apresentado será o de maximização de uma função utilidade suave (índice)

$$U = U(x, y) \quad (U_x, U_y > 0)$$

sujeita a

$$xP_x + yP_y = B$$

Condição de primeira ordem

A função de Lagrange desse modelo de otimização é

$$Z = U(x, y) + \lambda(B - xP_x - yP_y)$$

Como condição de primeira ordem, temos o seguinte conjunto de equações simultâneas:

$$\begin{aligned} Z_\lambda &= B - xP_x - yP_y = 0 \\ Z_x &= U_x - \lambda P_x = 0 \\ Z_y &= U_y - \lambda P_y = 0 \end{aligned} \quad (12.31)$$

Visto que as duas últimas equações são equivalentes a

$$\frac{U_x}{P_x} = \frac{U_y}{P_y} = \lambda \quad (12.31')$$

a condição de primeira ordem na verdade exige a satisfação de (12.31'), sujeita à restrição orçamentária – a primeira equação em (12.31). A equação de (12.31') é a mera e familiar proposição da teoria clássica do consumidor que diz que, para maximizar a utilidade, os consumidores devem alocar seus orçamentos de modo a igualar a razão entre a utilidade marginal e o preço para toda mercadoria. Especificamente, no ponto de equilíbrio ou ótimo, essas razões devem ter um valor comum λ^* . Como já sabemos, λ^* mede o efeito estático comparativo da constante de restrição sobre o valor ótimo da função objetivo. Por conseguinte, temos, no presente contexto,

$\lambda^* = (\partial U^*/\partial B)$; isto é, o valor ótimo do multiplicador de Lagrange pode ser interpretado como a *utilidade marginal do dinheiro* (verba orçamentária) quando a utilidade do consumidor é maximizada.

Se reescrevermos a condição em (12.31') na forma

$$\frac{U_x}{U_y} = \frac{P_x}{P_y} \quad (12.31'')$$

podemos dar uma interpretação alternativa à condição de primeira ordem em termos de curvas de indiferença.

Uma *curva de indiferença* é definida como o lugar geométrico das combinações de x e y que resultarão em um nível constante de U . Isso significa que devemos encontrar, sobre uma curva de indiferença,

$$dU = U_x dx + U_y dy = 0$$

o que implica que $dy/dx = -U_x/U_y$. De acordo com isso, se representarmos uma curva de indiferença no plano xy , como na Figura 12.7, sua inclinação, dy/dx , deve ser igual à negativa da razão de utilidade marginal U_x/U_y . (Posto que supusemos $U_x, U_y > 0$, a inclinação da curva de indiferença deve ser negativa.) Note que U_x/U_y , a negativa da inclinação da curva de indiferença, é denominada *taxa marginal de substituição* entre os dois bens.

E o significado de P_x/P_y ? Como veremos em breve, essa razão representa a negativa da inclinação do gráfico da restrição orçamentária. Essa restrição, $xP_x + yP_y = B$, pode ser escrita alternativamente como

$$y = \frac{B}{P_y} - \frac{P_x}{P_y} x$$

de modo que, quando representada graficamente no plano xy como na Figura 12.7, ela apareça como uma linha reta cuja inclinação é $-P_x/P_y$ (com interseção com o eixo vertical em B/P_y).

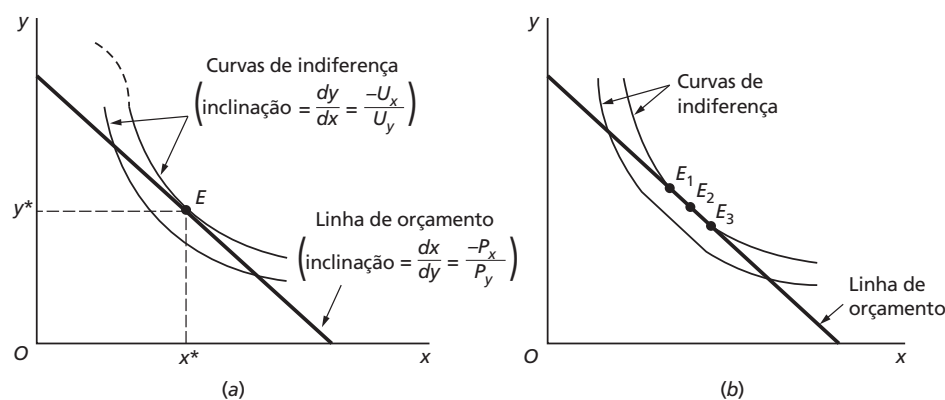
Sob essa perspectiva, a nova versão da condição de primeira ordem – (12.31'') mais a restrição orçamentária – revela que, para maximizar a utilidade, um consumidor deve alocar o orçamento de tal modo que a inclinação da reta de orçamento (sobre a qual o consumidor deve permanecer) é igual à inclinação de alguma curva de indiferença. Essa condição é satisfeita no ponto E na Figura 12.7a, onde a linha de orçamento é tangente a uma curva de indiferença.

Condição de segunda ordem

Se o hessiano aumentado no presente problema for positivo, isto é, se

$$|\bar{H}| = \begin{vmatrix} 0 & P_x & P_y \\ P_x & U_{xx} & U_{xy} \\ P_y & U_{yx} & U_{yy} \end{vmatrix} = 2P_x P_y U_{xy} - P_y^2 U_{xx} - P_x^2 U_{yy} > 0 \quad (12.32)$$

FIGURA 12.7



(com todas as derivadas calculadas nos valores críticos x^* e y^*), então o valor estacionário de U será, com certeza, um máximo. A presença das derivadas U_{xx} , U_{yy} e U_{xy} em (12.32) sugere claramente que satisfazer essa condição acarretaria certas restrições sobre a função utilidade e, portanto, sobre a forma das curvas de indiferença. Quais são essas restrições?

Considerando, em primeiro lugar, a forma das curvas de indiferença, podemos mostrar que um $|\bar{H}|$ positivo significa a convexidade estrita da curva de indiferença (de inclinação descendente) no ponto de tangência E . Exatamente como a inclinação descendente de uma curva de indiferença é garantida por uma $dy/dx (= -U_x/U_y)$ negativa, sua convexidade estrita estaria assegurada por uma d^2y/dx^2 positiva. Para obter a expressão para d^2y/dx^2 , podemos diferenciar $-U_x/U_y$ em relação a x ; mas, ao fazer isso, devemos ter em mente não apenas que ambas, U_x e U_y (por serem derivadas), são funções de x e y , mas também que, juntamente com uma curva de indiferença dada, y é, ela mesma, uma função de x . De acordo com isso, ambas, U_x e U_y , podem ser consideradas como funções apenas de x ; então, podemos obter uma derivada total

$$\frac{d^2y}{dx^2} = \frac{d}{dx} \left(-\frac{U_x}{U_y} \right) = -\frac{1}{U_y^2} \left(U_y \frac{dU_x}{dx} - U_x \frac{dU_y}{dx} \right) \quad (12.33)$$

Visto que x pode afetar U_x e U_y não somente diretamente, mas também indiretamente, através da intermediária y , temos

$$\frac{dU_x}{dx} = U_{xx} + U_{yx} \frac{dy}{dx} \quad \frac{dU_y}{dx} = U_{xy} + U_{yy} \frac{dy}{dx} \quad (12.34)$$

onde dy/dx refere-se à inclinação da curva de indiferença. Agora, no ponto de tangência E – o único ponto relevante para a discussão da condição de segunda ordem –, essa inclinação é idêntica à da restrição orçamentária; isto é, $dy/dx = -P_x/P_y$. Assim, podemos reescrever (12.34) como

$$\frac{dU_x}{dx} = U_{xx} + U_{yx} \frac{P_x}{P_y} \quad \frac{dU_y}{dx} = U_{xy} - U_{yy} \frac{P_x}{P_y} \quad (12.34')$$

Substituindo (12.34') em (12.33), utilizando a informação de que

$$U_x = \frac{U_y P_x}{P_y} \quad [\text{de (12.31'')}]$$

e então fatorando U_y/P_y^2 , podemos finalmente transformar (12.33) em

$$\frac{d^2y}{dx^2} = \frac{2P_x P_y U_{xy} - P_y^2 U_{xx} - P_x^2 U_{yy}}{U_y P_y^2} = \frac{|\bar{H}|}{U_y P_y^2} \quad (12.33')$$

Fica claro que, quando a condição suficiente de segunda ordem (12.32) é satisfeita, a derivada segunda em (12.33') é positiva e a curva de indiferença relevante é estritamente convexa no ponto de tangência. No presente contexto, também é verdadeiro que a convexidade estrita da curva de indiferença no ponto de tangência implica a satisfação da condição suficiente (12.32). Isso porque, dado que a inclinação das curvas de indiferença é negativa, sem nenhum ponto estacionário em lugar algum, a possibilidade de d^2y/dx^2 de valor zero sobre uma curva estritamente convexa está excluída. Assim, agora a convexidade estrita pode resultar somente em d^2y/dx^2 positiva e, portanto, em um $|\bar{H}|$ positivo, por (12.33').

Lembre-se, entretanto, que as derivadas em $|\bar{H}|$ devem ser calculadas somente nos valores críticos x^* e y^* . Assim, a convexidade estrita da curva de indiferença, como uma condição suficiente, é pertinente apenas ao ponto de tangência, e não é inconcebível que a curva contenha um segmento côncavo afastado do ponto E , como ilustrado pelo segmento de curva tracejado na Figura 12.7a. Por outro lado, se a função utilidade for reconhecidamente uma função suave, crescente e estritamente quase-côncava, então toda curva de indiferença será estritamente convexa em toda a sua extensão. A superfície dessa função utilidade é semelhante à que aparece na Figu-

ra 12.4b. Quando tal superfície é cortada por um plano paralelo ao plano xy , obtemos, para cada um desses cortes, uma seção transversal que, ao ser projetada sobre o plano xy , torna-se uma curva de indiferença estritamente convexa, de inclinação descendente. Nesse caso, não importa onde o ponto de tangência ocorra, a condição suficiente de segunda ordem sempre será satisfeita. Além disso, só pode existir um único ponto de tangência, um ponto que resulta no único nível máximo absoluto de utilidade que se pode obter com o orçamento linear dado. Esse resultado, é claro, está perfeitamente de acordo com o que afirma o losango à direita da Figura 12.6.

Temos repetido várias vezes que a condição suficiente de segunda ordem não é necessária. Agora vamos ilustrar a maximização da utilidade quando (12.32) não é válida. Suponha que, como ilustrado na Figura 12.7b, a curva de indiferença relevante contenha um segmento linear que coincide com uma porção da linha de orçamento. Então temos, claramente, vários máximos, visto que a condição de primeira ordem $U_x/U_y = P_x/P_y$ agora é satisfeita em todos os pontos sobre o segmento linear da curva de indiferença, incluindo E_1 , E_2 e E_3 . De fato, esses são máximos absolutos restritos. Mas, uma vez que, sobre um segmento de reta, d^2y/dx^2 é igual a zero, temos $|\bar{H}| = 0$ por (12.33'). Assim, a maximização é alcançada nesse caso, mesmo com a violação da condição suficiente de segunda ordem (12.32).

O fato de aparecer um segmento linear sobre a curva de indiferença sugere a presença de um segmento de plano inclinado sobre a superfície de utilidade. Isso ocorre quando a função utilidade é explicitamente quase-côncava em vez de estritamente quase-côncava. Como mostra a Figura 12.7b, os pontos E_1 , E_2 e E_3 , todos localizados sobre a mesma curva de indiferença (a mais alta que se pode obter) fornecem a mesma utilidade máxima absoluta sob a restrição orçamentária linear dada. Novamente com referência à Figura 12.6, notamos que esse resultado é perfeitamente consistente com a mensagem transmitida pelo losango à esquerda.

Análise estática comparativa

Em nosso modelo de consumidor, os preços P_x e P_y são exógenos, assim como a verba orçamentária B . Supondo que a condição suficiente de segunda ordem é satisfeita, podemos analisar as propriedades estáticas comparativas do modelo, com base na condição de primeira ordem (12.31), vista como um conjunto de equações $F^j = 0$ ($j = 1, 2, 3$), onde cada função F^j tem derivadas parciais contínuas. Como salientado em (12.19), o jacobiano de variáveis endógenas desse conjunto de equações deve ter o mesmo valor do hessiano aumentado; isto é, $|J| = |\bar{H}|$. Assim, quando a condição de segunda ordem (12.32) é satisfeita, $|J|$ deve ser positivo e não se anula no ótimo inicial. Por consequência, o teorema da função implícita pode ser aplicado e podemos expressar os valores ótimos das variáveis endógenas como funções implícitas das variáveis exógenas:

$$\begin{aligned}\lambda^* &= \lambda^*(P_x, P_y, B) \\ x^* &= x^*(P_x, P_y, B) \\ y^* &= y^*(P_x, P_y, B)\end{aligned}\tag{12.35}$$

Sabemos que essas funções possuem derivadas contínuas que fornecem informações estáticas comparativas. Em particular, as derivadas das duas últimas funções x^* e y^* , que descrevem o comportamento de demanda do consumidor, podem nos informar como ele reagirá a variações no preço e no orçamento. Para achar essas derivadas, contudo, precisamos, primeiro, converter (12.31) em um conjunto de identidades de equilíbrio da seguinte maneira:

$$\begin{aligned}B - x^*P_x - y^*P_y &\equiv 0 \\ U_x(x^*, y^*) - \lambda^*P_x &\equiv 0 \\ U_y(x^*, y^*) - \lambda^*P_y &\equiv 0\end{aligned}\tag{12.36}$$

Tomando a diferencial total de cada identidade por vez (permitindo que todas as variáveis se alterem) e notando que $U_{xy} = U_{yx}$ chegamos então ao sistema linear

$$\begin{aligned}
 -P_x dx^* - P_y dy^* &= x^* dP_x + y^* dP_y - dB \\
 -P_x d\lambda^* + U_{xx} dx^* + U_{xy} dy^* &= \lambda^* dP_x \\
 -P_y d\lambda^* + U_{yx} dx^* + U_{yy} dy^* &= \lambda^* dP_y
 \end{aligned} \tag{12.37}$$

Para estudar o efeito de uma variação no tamanho do orçamento (também denominado *renda* do consumidor), vamos supor $dP_x = dP_y = 0$, mas mantendo $dB \neq 0$. Então, após dividir todos os termos de (12.37) por dB , e interpretar cada razão de diferenciais como uma derivada parcial, podemos escrever a equação matricial[†]

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix} \begin{bmatrix} (\partial \lambda^* / \partial B) \\ (\partial x^* / \partial B) \\ (\partial y^* / \partial B) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \tag{12.38}$$

Como você pode verificar, o arranjo de elementos na matriz de coeficientes é exatamente o mesmo que apareceria no jacobiano $|J|$, que tem o mesmo valor do hessiano aumentado $|\bar{H}|$, embora este último tenha P_x e P_y (em vez de $-P_x$ e $-P_y$) na primeira linha e na primeira coluna. Pela regra de Cramer, podemos resolver para todas as três derivadas estáticas comparativas, mas limitaremos nossa atenção às duas seguintes:

$$\left(\frac{\partial x^*}{\partial B} \right) = \frac{1}{|J|} \begin{vmatrix} 0 & -1 & -P_y \\ -P_x & 0 & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} = \frac{1}{|J|} \begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix} \tag{12.39}$$

$$\left(\frac{\partial y^*}{\partial B} \right) = \frac{1}{|J|} \begin{vmatrix} 0 & -P_x & -1 \\ -P_x & U_{xx} & 0 \\ -P_y & U_{yx} & 0 \end{vmatrix} = \frac{-1}{|J|} \begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} \tag{12.40}$$

Pela condição de segunda ordem, $|J| = |\bar{H}|$ é positivo, assim como P_x e P_y . Infelizmente, na ausência de informação adicional sobre as grandezas relativas de P_x , P_y e U_{ij} , ainda não podemos averiguar os sinais dessas duas derivadas estáticas comparativas. Isso significa que, à medida que o orçamento (ou a renda) do consumidor aumenta, suas compras ótimas x^* e y^* podem aumentar ou diminuir. Caso, digamos, x^* diminua enquanto B aumenta, o produto x é denominado um *bem inferior*, em contraste a um *bem normal*.

Em seguida, podemos analisar o efeito de uma variação em P_x . Com $dP_y = dB = 0$ dessa vez, mas mantendo $dP_x \neq 0$, e então dividindo (12.37) por dP_x , obtemos uma outra equação matricial:

$$\begin{bmatrix} 0 & -P_x & -P_y \\ -P_x & U_{xx} & U_{xy} \\ -P_y & U_{yx} & U_{yy} \end{bmatrix} \begin{bmatrix} (\partial \lambda^* / \partial P_x) \\ (\partial x^* / \partial P_x) \\ (\partial y^* / \partial P_x) \end{bmatrix} = \begin{bmatrix} x^* \\ \lambda^* \\ 0 \end{bmatrix} \tag{12.41}$$

Disso emergem as derivadas estáticas comparativas:

$$\left(\frac{\partial x^*}{\partial P_x} \right) = \frac{1}{|J|} \begin{vmatrix} 0 & x^* & -P_y \\ -P_x & \lambda^* & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix}$$

[†] A equação matricial (12.38) também pode ser obtida diferenciando-se totalmente (12.36) em relação a B , tendo em mente as soluções implícitas em (12.35).

$$= \frac{-x^*}{|J|} \begin{vmatrix} -P_x & U_{xy} \\ -P_y & U_{yy} \end{vmatrix} + \frac{\lambda^*}{|J|} \begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix}$$

$$\equiv T_1 + T_2 \quad [T_i \text{ significa o } i\text{-ésimo termo}] \quad (12.42)$$

$$\left(\frac{\partial y^*}{\partial P_x} \right) = \frac{1}{|J|} \begin{bmatrix} 0 & -P_x & x^* \\ -P_x & U_{xx} & \lambda^* \\ -P_y & U_{yx} & 0 \end{bmatrix}$$

$$= \frac{x^*}{|J|} \begin{vmatrix} -P_x & U_{xx} \\ -P_y & U_{yx} \end{vmatrix} - \frac{\lambda^*}{|J|} \begin{vmatrix} 0 & -P_x \\ -P_y & U_{yx} \end{vmatrix}$$

$$\equiv T_3 + T_4 \quad (12.43)$$

Como interpretamos esses dois resultados? O primeiro, $(\partial x^* / \partial P_x)$, nos mostra como uma variação em P_x afeta a compra ótima de x ; assim, ele nos dá a base para o estudo da função demanda de nosso consumidor para x . Nesse sentido, há dois termos componentes. O primeiro termo, T_1 , pode ser reescrito, usando (12.39), como $-(\partial x^* / \partial B) x^*$. Sob essa perspectiva, T_1 parece ser uma medida do efeito de uma variação em B (orçamento ou renda) sobre a compra ótima x^* , com o próprio x^* servindo como fator de ponderação ou peso. Todavia, visto que essa derivada obviamente diz respeito a uma variação no preço, T_1 deve ser interpretado como o *efeito sobre a renda* causado por uma *variação no preço*. À medida que P_x aumenta, o declínio na renda real do consumidor produzirá sobre x^* um efeito semelhante a uma redução propriamente dita em B ; daí a utilização do termo $-(\partial x^* / \partial B)$. Portanto, é de se compreender que, quanto mais proeminente for o lugar ocupado pela mercadoria x no orçamento total, maior será seu efeito sobre a renda – e, portanto, a presença do fator de ponderação x^* em T_1 . Essa interpretação pode ser demonstrada de modo mais formal expressando a perda efetiva de renda do consumidor pela diferencial $dB = -x^* dP_x$. Então, temos

$$x^* = \frac{dB}{dP_x} \quad (12.44)$$

e

$$T_1 = - \left(\frac{\partial x^*}{\partial B} \right) x^* = \left(\frac{\partial x^*}{\partial B} \right) \frac{dB}{dP_x}$$

que mostra que T_1 é a medida do efeito de dP_x sobre x^* por meio de B , isto é, o efeito da renda.

Agora, se compensarmos o consumidor pela perda efetiva de renda com um pagamento em fundos numericamente igual a dB , então, por causa da neutralização do efeito renda, o componente remanescente na derivada estática comparativa $(\partial x^* / \partial P_x)$, a saber, T_2 , medirá a variação de x^* devida inteiramente à substituição de uma mercadoria por outra induzida pelo preço, isto é, o *efeito de substituição* da variação de P_x . Para ver isso com maior clareza, vamos voltar a (12.37) e considerar como a compensação de renda modificará a situação. Quando estudamos apenas o efeito de dP_x (com $dP_y = dB = 0$), a primeira equação em (12.37) pode ser escrita como $-P_x dx^* - P_y dy^* = x^* dP_x$. Visto que a indicação da perda de renda efetiva do consumidor está na expressão $x^* dP_x$ (que, a propósito, aparece somente na primeira equação), compensar o consumidor significa igualar esse termo a zero. Então, o vetor de constantes em (12.41) deve ser trocado de

$$\begin{bmatrix} x^* \\ \lambda^* \\ 0 \end{bmatrix} \text{ para } \begin{bmatrix} 0 \\ \lambda^* \\ 0 \end{bmatrix}, \text{ e a versão de renda compensada da derivada } (\partial x^* / \partial P_x) \text{ será}$$

$$\left(\frac{\partial x^*}{\partial P_x} \right)_{\text{compensada}} = \frac{1}{|J|} \begin{vmatrix} 0 & 0 & -P_y \\ -P_x & \lambda^* & U_{xy} \\ -P_y & 0 & U_{yy} \end{vmatrix} = \frac{\lambda^*}{|J|} \begin{vmatrix} 0 & -P_y \\ -P_y & U_{yy} \end{vmatrix} = T_2$$

Portanto, podemos expressar (12.42) na forma

$$\left(\frac{\partial x^*}{\partial P_x} \right) = T_1 + T_2 = \underbrace{- \left(\frac{\partial x^*}{\partial B} \right) x^*}_{\text{efeito renda}} + \underbrace{\left(\frac{\partial x^*}{\partial P_x} \right)_{\text{compensada}}}_{\text{efeito substituição}} \quad (12.42')$$

Esse resultado, que decompõe a derivada estática comparativa $(\partial x^* / \partial P_x)$ em dois componentes, um efeito renda e um efeito substituição, é a versão de dois bens da denominada equação de Slutsky.

O que podemos dizer sobre o sinal de $(\partial x^* / \partial P_x)$? O efeito substituição T_2 é claramente negativo, porque $|J| > 0$ e $\lambda^* > 0$ [veja (12.31')]. O efeito renda T_1 , por outro lado, tem sinal indeterminado de acordo com (12.39). Se for negativo, ele reforça T_2 ; nesse caso, um aumento em P_x tem de reduzir a compra de x , e a inclinação da curva de demanda do consumidor que maximiza a utilidade será negativa. Se for positivo, mas de grandeza relativamente pequena, ele diluirá o efeito substituição, embora o resultado geral ainda seja uma curva de demanda com inclinação descendente. Mas, no caso de T_1 ser positivo e prevalecer sobre T_2 (tal como acontece quando x^* é um item significativo no orçamento do consumidor, fornecendo um fator de ponderação poderoso), então um aumento em P_x na verdade levará a uma compra *maior* de x , uma situação de demanda característica dos denominados *bens de Giffen*. Normalmente, é claro que esperaríamos que $(\partial x^* / \partial P_x)$ fosse negativa.

Por fim, vamos examinar a derivada estática comparativa em (12.43), $(\partial y^* / \partial P_x) = T_3 + T_4$, que tem a ver com o *efeito cruzado* de uma variação no preço de x sobre a compra ótima de y . O termo T_3 é notavelmente parecido com o termo T_1 e, mais uma vez, é interpretado como um efeito renda.[†] Note que o fator de ponderação aqui é, novamente, x^* (em vez de y^*); isso porque estamos estudando o efeito de uma variação de P_x sobre a renda efetiva, cuja magnitude depende da importância relativa de x^* (e não y^*) no orçamento do consumidor. Naturalmente, o termo remanescente, T_4 , é, novamente, uma medida do efeito substituição.

O sinal de T_3 , de acordo com (12.40), dependente de fatores como U_{xx} , U_{yx} etc., e é indeterminado sem restrições adicionais no modelo. Todavia, o efeito substituição T_4 com certeza será positivo em nosso modelo, visto que λ^* , P_x , P_y e $|J|$ são todos positivos. Isso significa que, a menos que seja mais do que deslocado por um efeito renda negativo, um aumento no preço de x sempre aumentará a compra de y em nosso modelo de duas mercadorias. Em outras palavras, no contexto do presente modelo, no qual o consumidor pode escolher somente entre dois bens, esses bens devem guardar uma relação entre eles como substitutos.

Mesmo que a análise precedente se refira aos efeitos de uma variação em P_x , nossos resultados são adaptáveis de imediato ao caso de uma variação em P_y . Acontece que nosso modelo é tal que as posições ocupadas pelas variáveis x e y são perfeitamente simétricas. Assim, para inferir os efeitos de uma variação em P_y , basta permutar os papéis de x e y nos resultados já obtidos.

[†] Se você precisar de uma dose maior de segurança de que T_3 representa o efeito renda, pode usar (12.40) e (12.44) para escrever

$$T_3 = - \left(\frac{\partial y^*}{\partial B} \right) x^* = \left(\frac{\partial y^*}{\partial B} \right) \frac{dB}{dP_x}$$

Assim, T_3 é o efeito de uma variação P_x sobre y^* por meio do fator renda B .

Variações proporcionais em preços e renda

Também é de interesse indagar como x^* e y^* serão afetados quando todos os três parâmetros P_x , P_y e B variarem na mesma proporção. Essa questão ainda está no âmbito da estática comparativa, mas, diferentemente da análise precedente, o que investigamos agora envolve a mudança simultânea de todos os parâmetros.

Quando ambos os preços são aumentados, juntamente com a renda, pelo mesmo múltiplo j , cada termo na restrição orçamentária aumentará j vezes, e teremos:

$$jB - jxP_x - jyP_y = 0$$

Entretanto, considerando que o fator comum j pode ser cancelado, essa nova restrição é, de fato, idêntica à antiga. A função utilidade, além disso, é independente desses parâmetros. Por consequência, os antigos níveis de equilíbrio de x e y continuarão a prevalecer; isto é, a posição de equilíbrio do consumidor em nosso modelo é invariante com mudanças proporcionais *iguais* em todos os preços e na renda. Assim, no presente modelo, o consumidor é considerado livre de qualquer “ilusão monetária”.

Simbolicamente, essa situação pode ser descrita pelas equações

$$\begin{aligned} x^*(P_x, P_y, B) &= x^*(jP_x, jP_y, jB) \\ y^*(P_x, P_y, B) &= y^*(jP_x, jP_y, jB) \end{aligned}$$

As funções x^* e y^* , com a propriedade de *invariância* que acabamos de citar, não são funções comuns; são exemplos de uma classe especial de funções conhecidas como *funções homogêneas*, as quais têm interessantes aplicações econômicas. Portanto, examinaremos essas funções na Seção 12.6.

EXERCÍCIO 12.5

- Dado $U = (x + 2)(y + 1)$ e $P_x = 4$, $P_y = 6$ e $B = 130$:
 - Escreva a função de Lagrange.
 - Encontre os níveis ótimos de compra x^* e y^* .
 - A condição suficiente de segunda ordem para um máximo está satisfeita?
 - A resposta em (b) dá alguma informação estática comparativa?
- Suponha que $U = (x + 2)(y + 1)$ mas, desta vez, não atribua valores numéricos específicos aos parâmetros de preço e renda.
 - Escreva a função de Lagrange.
 - Encontre x^* , y^* e λ^* em termos dos parâmetros P_x , P_y e B .
 - Verifique a condição suficiente de segunda ordem para um máximo.
 - Supondo $P_x = 4$, $P_y = 6$ e $B = 130$, verifique a validade de sua resposta ao Problema 1.
- Sua solução (x^* e y^*) no Problema 2 pode fornecer alguma informação estática comparativa? Encontre todas as derivadas estáticas comparativas que puder, encontre seus sinais e interprete seus significados econômicos.
- A partir da função utilidade $U = (x + 2)(y + 1)$ e da restrição $xP_x + yP_y = B$ do Problema 2, já encontramos as U_{ij} e $|H|$, bem como x^* e λ^* . Além disso, lembramos que $|J| = |H|$.
 - Substitua essas expressões em (12.39) e (12.40) para encontrar $(\partial x^* / \partial B)$ e $(\partial y^* / \partial B)$.
 - Substitua em (12.42) e (12.43) para encontrar $(\partial x^* / \partial P_x)$ e $(\partial y^* / \partial P_x)$. Esses resultados estão de acordo com os obtidos no Problema 3?
- Comente a validade da afirmação: “Se a derivada $(\partial x^* / \partial P_x)$ for negativa, então x não pode representar um bem inferior”.

6. Ao estudar o efeito isolado de dP_x a primeira equação em (12.37) se reduz a $-P_x dx^* - P_y dy^* = x^* dP_x$ e, quando compensamos o consumidor pela perda de renda efetiva descartando o termo $x^* dP_x$, a equação se torna $-P_x dx^* - P_y dy^* = 0$. Mostre que este último resultado pode ser obtido alternativamente por um procedimento de compensação pelo qual tentamos manter inalterado o nível ótimo de utilidade do consumidor, U^* , (em vez da renda efetiva), de modo que o termo T_2 pode ser interpretado alternativamente como $(\partial x^* / \partial P_x)_{U^*=\text{constante}}$. [Sugestão: Utilize (12.31'').]
7. (a) A premissa de utilidade marginal decrescente para os bens x e y implica curvas de indiferença estritamente convexas?
(b) A premissa de convexidade estrita nas curvas de indiferença implica utilidade marginal decrescente para os bens x e y ?

12.6 Funções homogêneas

Diz-se que uma função é homogênea de grau r se a multiplicação de cada uma de suas variáveis independentes por uma constante j alterar o valor da função na proporção de j^r , isto é, se

$$f(jx_1, \dots, jx_n) = j^r f(x_1, \dots, x_n)$$

Em geral, j pode assumir qualquer valor. Todavia, para que a equação precedente faça sentido, $(jx_1, \dots, jx_n)$ não deve estar fora do domínio da função f . Por essa razão, em aplicações econômicas, a constante j é usualmente considerada positiva, pois a maioria das variáveis econômicas não admite valores negativos.

EXEMPLO 1 Dada a função $f(x, y, w) = x/y + 2w/3x$, se multiplicarmos cada variável por j , obteremos

$$f(jx, jy, jw) = \frac{(jx)}{(jy)} + \frac{2(jw)}{3(jx)} = \frac{x}{y} + \frac{2w}{3x} = f(x, y, w) = j^0 f(x, y, w)$$

Neste exemplo particular, o valor da função não será absolutamente afetado por variações proporcionais iguais em todas as variáveis independentes; ou, em outras palavras, o valor da função é trocado por um múltiplo de $j^0 (= 1)$. Isso torna a função f uma função homogênea de grau zero.

Você observará que as funções x^* e y^* citadas ao final da Seção 12.5 são ambas homogêneas de grau zero.

EXEMPLO 2 Quando multiplicamos cada variável na função

$$g(x, y, w) = \frac{x^2}{y} + \frac{2w^2}{x}$$

por j , obtemos

$$g(jx, jy, jw) = \frac{(jx)^2}{(jy)} + \frac{2(jw)^2}{(jx)} = j \left(\frac{x^2}{y} + \frac{2w^2}{x} \right) = jg(x, y, w)$$

A função g é homogênea de grau um (ou de primeiro grau); a multiplicação de cada variável por j também alterará o valor da função exatamente j vezes.

EXEMPLO 3 Agora, considere a função $h(x, y, w) = 2x^2 + 3yw - w^2$. Uma multiplicação semelhante nos dará, desta vez

$$h(jx, jy, jw) = 2(jx)^2 + 3(jy)(jw) - (jw)^2 = j^2 h(x, y, w)$$

Assim, a função h é homogênea de grau dois; neste caso, dobrar todas as variáveis, por exemplo, quadruplicará o valor da função.

Homogeneidade linear

Quando discutimos funções produção, fazemos amplo uso de funções de primeiro grau. Estas funções são geralmente denominadas funções *linearmente homogêneas*, sendo que o advérbio *linearmente* qualifica o adjetivo *homogênea*. Alguns autores, entretanto, parecem preferir a terminologia de certo modo enganosa, funções homogêneas *lineares*, ou até mesmo funções *lineares e homogêneas*, o que tende a transmitir, erroneamente, a impressão de que as próprias funções são lineares. Com base na função g do Exemplo 2, sabemos que uma função que é homogênea de primeiro grau *não é necessariamente* linear, em si. Por conseguinte, você deve evitar utilizar os termos “funções lineares homogêneas” e “funções lineares e homogêneas”, a menos que, é claro, as funções em questão sejam, de fato, lineares. Note, contudo, que não é incorreto falar de “homogeneidade linear”, no sentido de homogeneidade de grau um porque qualificar um substantivo (homogeneidade) exige a utilização de um adjetivo (linear).

Visto que a área primária de aplicação de funções linearmente homogêneas é a teoria da produção, vamos adotar, como estrutura para nossa discussão, uma função produção da forma, digamos,

$$Q = f(K, L) \quad (12.45)$$

Seja aplicada no nível *micro* ou no nível *macro*, a premissa matemática de homogeneidade linear equivaleria à premissa econômica de retornos constantes de escala, porque homogeneidade linear significa que aumentar todos os insumos (variáveis independentes) j vezes sempre elevará a produção resultante (valor da função) exatamente j vezes.

Quais propriedades singulares caracterizam essa função produção linearmente homogênea?

Propriedade I Dada a função produção linearmente homogênea $Q = f(K, L)$, o produto físico médio do trabalho (APP_L) e do capital (APP_K) podem ser expressos como funções da razão capital-trabalho, $k \equiv K/L$, apenas.

Para provar isso, multiplicamos cada variável independente em (12.45) por um fator $j = 1/L$. Em virtude da homogeneidade linear, isso mudará a forma do produto de Q para $jQ = Q/L$. O lado direito de (12.45) se tornará, correspondentemente,

$$f\left(\frac{K}{L}, \frac{L}{L}\right) = f\left(\frac{K}{L}, 1\right) = f(k, 1)$$

Visto que as variáveis K e L na função original deverão ser substituídas (sempre que aparecerem) por k e 1, respectivamente, o lado direito se torna, na verdade, uma função da razão isolada capital-trabalho k digamos, $\phi(k)$, que é uma função que tem um único argumento, k , mesmo que, na verdade, duas variáveis independentes K e L estejam envolvidas naquele argumento. Igualando os dois lados, temos

$$APP_L \equiv \frac{Q}{L} = \phi(k) \quad (12.46)$$

Constatamos, então, que a expressão para APP_K é

$$APP_K \equiv \frac{Q}{K} = \frac{Q}{L} \frac{L}{K} = \frac{\phi(k)}{k} \quad (12.47)$$

Uma vez que ambos os produtos médios dependem apenas de k , a homogeneidade linear implica que, contanto que a razão K/L seja mantida constante (sejam quais forem os níveis absolutos de K e L), os produtos médios também serão constantes. Por conseguinte, conquanto a fun-

ção produção seja homogênea de grau um, ambos, APP_L e APP_K , são homogêneos de grau zero nas variáveis K e L , já que variações proporcionais iguais em K e L (mantendo um k constante) não alterarão as grandezas dos produtos médios.

Propriedade II Dada uma função produção linearmente homogênea $Q = f(K, L)$, os produtos físicos marginais MPP_L e MPP_K podem ser expressos como funções apenas de k .

Para achar os produtos marginais, em primeiro lugar, escrevemos o produto total como

$$Q = L\phi(k) \quad [\text{por (12.46)}] \quad (12.45')$$

e então diferenciamos Q em relação a K e L . Para essa finalidade, veremos que os dois resultados preliminares seguintes são úteis:

$$\frac{\partial k}{\partial K} = \frac{\partial}{\partial K} \left(\frac{K}{L} \right) = \frac{1}{L} \quad \frac{\partial k}{\partial L} = \frac{\partial}{\partial L} \left(\frac{K}{L} \right) = \frac{-K}{L^2} \quad (12.48)$$

Os resultados da diferenciação são

$$\begin{aligned} MPP_K &\equiv \frac{\partial Q}{\partial K} = \frac{\partial}{\partial K} [L\phi(k)] \\ &= L \frac{\partial \phi(k)}{\partial K} = L \frac{d\phi(k)}{dk} \frac{\partial k}{\partial K} \quad [\text{regra da cadeia}] \\ &= L\phi'(k) \left(\frac{1}{L} \right) = \phi'(k) \quad [\text{por (12.48)}] \end{aligned} \quad (12.49)$$

$$\begin{aligned} MPP_L &\equiv \frac{\partial Q}{\partial L} = \frac{\partial}{\partial L} [L\phi(k)] \\ &= \phi(k) + L \frac{\partial \phi(k)}{\partial L} \quad [\text{regra do produto}] \\ &= \phi(k) + L\phi'(k) \frac{\partial k}{\partial L} \quad [\text{regra da cadeia}] \\ &= \phi(k) + L\phi'(k) \frac{-K}{L^2} \quad [\text{por (12.48)}] \\ &= \phi(k) - k\phi'(k) \end{aligned} \quad (12.50)$$

os quais mostram, de fato, que MPP_K e MPP_L são funções apenas de k .

Como são produtos médios, os produtos marginais permanecerão os mesmos contanto que a razão capital-trabalho seja mantida constante; eles são homogêneos de grau zero nas variáveis K e L .

Propriedade III (teorema de Euler) Se $Q = f(K, L)$ for linearmente homogênea, então

$$K \frac{\partial Q}{\partial K} + L \frac{\partial Q}{\partial L} \equiv Q$$

DEMONSTRAÇÃO

$$\begin{aligned} K \frac{\partial Q}{\partial K} + L \frac{\partial Q}{\partial L} &= K\phi'(k) + L[\phi(k) - k\phi'(k)] \quad [\text{by (12.49), (12.50)}] \\ &= K\phi'(k) + L\phi(k) - K\phi'(k) \quad [k \equiv K/L] \\ &= L\phi(k) = Q \quad [\text{por (12.45')}] \end{aligned}$$

Note que esse resultado é válido para *quaisquer* valores de K e L ; é por isso que a propriedade pode ser escrita como uma identidade. Essa propriedade diz que o valor de uma função linearmente homogênea sempre pode ser expresso como uma soma de termos, sendo cada qual o produto de uma das variáveis independentes pela derivada parcial de primeira ordem em relação àquela variável, independentemente dos níveis dos dois insumos realmente empregados.

Tome o cuidado, entretanto, de distinguir entre a identidade $K \frac{\partial Q}{\partial K} + L \frac{\partial Q}{\partial L} \equiv Q$ [teorema de Euler, que se aplica somente ao caso de retornos constantes de escala de $Q = f(K, L)$] e a equação $dQ = \frac{\partial Q}{\partial K} dK + \frac{\partial Q}{\partial L} dL$ [diferencial total de Q , para *qualquer* função $Q = f(K, L)$].

Em termos econômicos, essa propriedade significa que, sob condições de retornos constantes de escala, se cada fator de insumo for pago por uma quantia equivalente a seu produto marginal, o produto total será exatamente exaurido pelas participações distributivas de todos os fatores de insumo ou, o que é equivalente, o lucro econômico puro será igual a zero. Visto que essa situação descreve o equilíbrio de longo prazo sob competição pura, considerava-se que somente funções produção linearmente homogêneas fariam sentido em economia. Evidentemente, não é esse o caso. O lucro econômico zero no equilíbrio de longo prazo é causado pelas forças da concorrência por meio de entradas e saídas de empresas, independentemente da natureza específica das funções produção que estão realmente prevalecendo. Assim, não é obrigatório ter uma função produção que garanta a exaustão de produto para todos e quaisquer pares (K, L) . Além do mais, quando existe concorrência imperfeita nos mercados dos fatores, a remuneração dos fatores pode não ser igual aos produtos marginais e, por consequência, o teorema de Euler se torna irrelevante para o quadro da distribuição. Contudo, muitas vezes é conveniente trabalhar com funções produção linearmente homogêneas por causa das elegantes propriedades matemáticas que elas reconhecidamente possuem.

A função produção de Cobb-Douglas

Uma função produção específica amplamente usada na análise econômica (citada anteriormente na Seção 11.6, Exemplo 5) é a *função produção de Cobb-Douglas*:

$$Q = AK^\alpha L^{1-\alpha} \quad (12.51)$$

onde A é uma constante positiva e α é uma fração positiva. Aqui consideraremos, em primeiro lugar, uma versão generalizada dessa função, a saber,

$$Q = AK^\alpha L^\beta \quad (12.52)$$

onde β é uma outra fração positiva que pode ou não ser igual a $1 - \alpha$. Algumas das características importantes dessa função são: (1) ela é homogênea de grau $(\alpha + \beta)$; (2) no caso especial de $\alpha + \beta = 1$, ela é linearmente homogênea; (3) suas isoquantas têm inclinações negativas em toda a sua extensão e são estritamente convexas para valores positivos de K e L ; e (4) ela é estritamente quase-côncava para K e L positivos.

Sua homogeneidade pode ser facilmente percebida pelo fato de que, trocando K e L por jK e jL , respectivamente, o resultado da produção mudará para

$$A(jK)^\alpha (jL)^\beta = j^{\alpha+\beta} (AK^\alpha L^\beta) = j^{\alpha+\beta} Q$$

Isto é, a função é homogênea de grau $(\alpha + \beta)$. No caso de $\alpha + \beta = 1$, haverá retornos constantes de escala, porque a função será linearmente homogênea. (Note, entretanto, que essa função *não* é linear! Assim, ficaria confuso se nos referíssemos a ela como uma função “linear homogênea” ou “linear e homogênea”.) O fato de suas isoquantas terem inclinações negativas e convexidade estrita pode ser verificado pelos sinais das derivadas dK/dL e d^2K/dL^2 (ou pelos sinais de dL/dK e d^2L/dK^2). Para qualquer resultado de produção positivo Q_0 , (12.52) pode ser escrita como

$$AK^\alpha L^\beta = Q_0 \quad (A, K, L, Q_0 > 0)$$

Tomando o logaritmo natural de ambos os lados e transpondo, constatamos que

$$\ln A + \alpha \ln K + \beta \ln L - \ln Q_0 = 0$$

que define, implicitamente, K como uma função de L .[†] Portanto, pela regra da função implícita e pela regra do logaritmo, temos

$$\frac{dK}{dL} = - \frac{\partial F / \partial L}{\partial F / \partial K} = - \frac{(\beta / L)}{(\alpha / K)} = - \frac{\beta K}{\alpha L} < 0$$

Então, segue-se que

$$\frac{d^2 K}{dL^2} = \frac{d}{dL} \left(- \frac{\beta K}{\alpha L} \right) = - \frac{\beta}{\alpha} \frac{d}{dL} \left(\frac{K}{L} \right) = - \frac{\beta}{\alpha} \frac{1}{L^2} \left(L \frac{dK}{dL} - K \right) > 0$$

Os sinais das derivadas estabelecem que a inclinação da isoquanta (qualquer isoquanta) é descendente em toda a sua extensão e que ela é estritamente convexa no plano LK para valores positivos de K e L . Isso, é claro, é o que esperamos de uma função que é estritamente quase-côncava para K e L positivos. Se quiser verificar a característica de quase-concavidade dessa função, veja o Exemplo 5 da Seção 12.4, onde foi discutida uma função similar.

Agora vamos examinar o caso $\alpha + \beta = 1$ (a função de Cobb-Douglas propriamente dita), para verificar as três propriedades de homogeneidade linear citadas anteriormente. Em primeiro lugar, o produto total nesse caso especial pode ser expresso como

$$Q = AK^\alpha L^{1-\alpha} = A \left(\frac{K}{L} \right)^\alpha L = LAK^\alpha \quad (12.51')$$

onde a expressão AK^α é uma versão específica da expressão geral $\phi(k)$ usada anteriormente. Portanto, os produtos médios são

$$APP_L = \frac{Q}{L} AK^\alpha \quad (12.53)$$

$$APP_K = \frac{Q}{K} = \frac{Q}{L} \frac{L}{K} = \frac{AK^\alpha}{k} = AK^{\alpha-1}$$

ambos os quais agora são funções apenas de k .

Em segundo lugar, a diferenciação de $Q = AK^\alpha L^{1-\alpha}$ resulta nos produtos marginais:

$$\frac{\partial Q}{\partial K} = A\alpha K^{\alpha-1} L^{-(\alpha-1)} = A\alpha \left[\frac{K}{L} \right]^{\alpha-1} = A\alpha k^{\alpha-1} \quad (12.54)$$

$$\frac{\partial Q}{\partial L} = AK^\alpha (1-\alpha) L^{-\alpha} = A(1-\alpha) \left[\frac{K}{L} \right]^\alpha = A(1-\alpha) k^\alpha$$

que também são funções apenas de k .

Por fim, podemos verificar o teorema de Euler usando (12.54) como se segue:

$$K \frac{\partial Q}{\partial K} + L \frac{\partial Q}{\partial L} = K A \alpha k^{\alpha-1} + L A (1-\alpha) k^\alpha$$

[†] As condições do teorema da função implícita são satisfeitas porque F (a expressão do lado esquerdo) tem derivadas parciais contínuas e porque $\partial F / \partial K = \alpha K \neq 0$ para valores positivos de K .

$$\begin{aligned}
 &= L A k^{\alpha} \left(\frac{K\alpha}{Lk} + 1 - \alpha \right) \\
 &= L A k^{\alpha} (\alpha + 1 - \alpha) = L A k^{\alpha} = Q \quad [\text{por (12.51')}]
 \end{aligned}$$

Interessantes significados econômicos podem ser atribuídos aos expoentes α e $(1 - \alpha)$ na função produção de Cobb-Douglas linearmente homogênea. Se admitirmos que cada insumo é pago pela quantia equivalente a seu produto marginal, a participação relativa do produto total originada do capital será

$$\frac{K(\partial Q / \partial K)}{Q} = \frac{K A \alpha k^{\alpha-1}}{L A k^{\alpha}} = \alpha$$

De modo semelhante, a participação relativa do trabalho será

$$\frac{L(\partial Q / \partial L)}{Q} = \frac{L A (1 - \alpha) k^{\alpha}}{L A k^{\alpha}} = 1 - \alpha$$

Assim, o expoente de cada variável insumo indica a participação relativa daquele insumo no produto total. De outra perspectiva, também podemos interpretar o expoente de cada variável insumo como a elasticidade parcial do produto em relação àquele insumo. Isso porque a expressão da participação do capital que acabamos de dar é equivalente à expressão $\frac{\partial Q / \partial K}{Q / K} \equiv \varepsilon_{QK}$ e, de

modo semelhante, a expressão da participação do trabalho que acabamos de dar é precisamente a de ε_{QL} .

E o significado da constante A ? Para valores dados de K e L , a grandeza de A afetará proporcionalmente o nível de Q . Portanto, A pode ser considerado como um *parâmetro de eficiência*, isto é, como um indicador do estado da tecnologia.

Extensões dos resultados

Discutimos homogeneidade linear no contexto específico de funções produção, mas as propriedades citadas são igualmente válidas em outros contextos, contanto que as variáveis K , L e Q sejam adequadamente reinterpretadas.

Além do mais, é possível estender nossos resultados ao caso de mais de duas variáveis. Com uma função linearmente homogênea

$$y = f(x_1, x_2, \dots, x_n)$$

novamente podemos dividir cada variável por x_1 (isto é, multiplicar por $1/x_1$) e obter o resultado

$$y = x_1 \phi \left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \dots, \frac{x_n}{x_1} \right) \quad [\text{homogeneidade de grau 1}]$$

que é comparável a (12.45'). Além do mais, o teorema de Euler é estendido com facilidade para a forma

$$\sum_{i=1}^n x_i f_i \equiv y \quad [\text{teorema de Euler}]$$

onde as derivadas parciais da função original f (a saber, f_i) novamente são homogêneas de grau zero nas variáveis x_i , como acontece no caso de duas variáveis.

As extensões precedentes também podem, de fato, ser generalizadas com relativa facilidade para uma função homogênea de grau r . Em primeiro lugar, pela definição de homogeneidade, podemos escrever, no presente caso

$$y = x_1^r \phi\left(\frac{x_2}{x_1}, \frac{x_3}{x_1}, \dots, \frac{x_n}{x_1}\right) \quad [\text{homogeneidade de grau } r]$$

A versão modificada do teorema de Euler aparecerá agora na forma

$$\sum_{i=1}^n x_i f_i \equiv ry \quad [\text{teorema de Euler}]$$

onde uma constante multiplicativa r foi anexada à variável dependente y à direita. E, finalmente, as derivadas parciais da função original f , as funções f_i , serão todas homogêneas de grau $(r - 1)$ nas variáveis x_i . Assim, você pode ver que o caso da homogeneidade linear é um mero caso especial dela, no qual $r = 1$.

EXERCÍCIO 12.6

- Determine se as seguintes funções são homogêneas. Se forem, quais são seus graus?
 - $f(x, y) = \sqrt{xy}$
 - $f(x, y) = (x^2 - y^2)^{1/2}$
 - $f(x, y) = x^3 - xy + y^3$
 - $f(x, y) = 2x + y + 3\sqrt{xy}$
 - $f(x, y, w) = \frac{xy^2}{w} + 2xw$
 - $f(x, y, w) = x^4 - 5yw^3$
- Mostre que a função (12.45) pode ser expressa alternativamente como $Q = K\psi\left(\frac{L}{K}\right)$ em vez de $Q = L\phi\left(\frac{K}{L}\right)$.
- Deduzo do teorema de Euler que, com retornos constantes de escala:
 - Quando $MPP_K = 0$, APP_L é igual a MPP_L .
 - Quando $MPP_L = 0$, APP_K é igual a MPP_K .
- Com base em (12.46) até (12.50), verifique se as seguintes afirmações são válidas sob condições de retornos constantes de escala:
 - Uma curva APP_L pode ser traçada tendo $k (= K/L)$ como a variável independente (no eixo horizontal).
 - MPP_K é medido pela inclinação daquela curva APP_L .
 - APP_K é medido pela inclinação do vetor raio em relação à curva APP_L .
 - $MPP_L = APP_L - k(MPP_K) = APP_L - k$ (inclinação de APP_L).
- Use (12.53) e (12.54) para verificar que as relações descritas no Problema 4b, c e d são obedecidas pela função produção de Cobb-Douglas.
- Dada a função produção $Q = AK^\alpha L^\beta$, mostre que:
 - $\alpha + \beta > 1$ implica retornos crescentes de escala.
 - $\alpha + \beta < 1$ implica retornos decrescentes de escala.
 - α e β são, respectivamente, as elasticidades parciais de produto relativas aos insumos capital e trabalho.
- Seja o resultado da produção uma função de três insumos: $Q = AK^a L^b N^c$.
 - Essa função é homogênea? Se for, qual é seu grau?
 - Sob qual condição haveria retornos constantes de escala? E retornos crescentes de escala?
 - Encontre a participação de produto para o insumo N , se ele for pago pela quantia equivalente a seu produto marginal.
- Suponha que a função produção $Q = g(K, L)$ é homogênea de grau 2.
 - Escreva uma equação para expressar a homogeneidade de segundo grau dessa função.
 - Encontre uma expressão para Q em termos de $\phi(k)$, com as características de (12.45').
 - Encontre a função MPP_K . MPP_K ainda é uma função apenas de k , como no caso da homogeneidade linear?
 - A função MPP_K é homogênea em K e L ? Se for, qual é seu grau?

12.7 Combinação de insumos de custo mínimo

Como um outro exemplo de otimização restrita, vamos discutir o problema de achar a combinação de insumos de custo mínimo para a produção de um nível especificado de produto Q_0 que represente, digamos, o pedido especial de um cliente. Aqui, trabalharemos com uma função produção geral; mais adiante, contudo, faremos referência a funções de produção homogêneas.

Condição de primeira ordem

Supondo uma função produção suave com duas variáveis insumo, $Q = Q(a, b)$, onde $Q_a, Q_b > 0$, e supondo que o preço de ambos os insumos seja exógeno (embora, mais uma vez omitindo o índice zero), podemos formular o problema como um problema de minimização de custo

$$C = aP_a + bP_b$$

sujeito à restrição de produto

$$Q(a, b) = Q_0$$

Por conseguinte, a função de Lagrange é

$$Z = aP_a + bP_b + \mu[Q_0 - Q(a, b)]$$

Para satisfazer a condição de primeira ordem para um mínimo C , os níveis de insumo (as variáveis de escolha) devem satisfazer as seguintes equações simultâneas:

$$Z_\mu = Q_0 - Q(a, b) = 0$$

$$Z_a = P_a - \mu Q_a = 0$$

$$Z_b = P_b - \mu Q_b = 0$$

A primeira equação nesse conjunto é a mera restrição enunciada de outra maneira, e as duas últimas implicam a condição

$$\frac{P_a}{Q_a} = \frac{P_b}{Q_b} = \mu \quad (12.55)$$

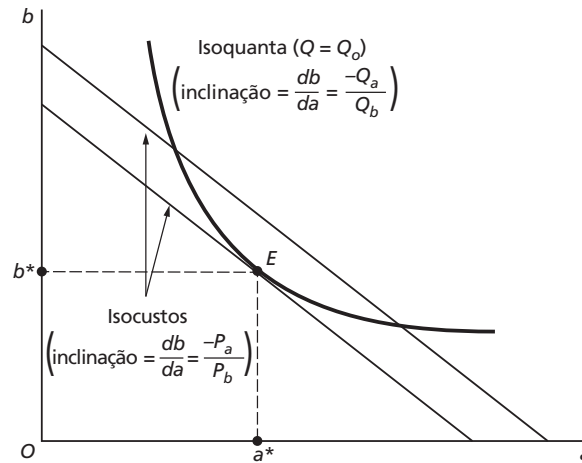
No ponto da combinação ótima de insumos, a razão entre preço do insumo e produto marginal deve ser a mesma para cada insumo. Visto que essa razão mede a quantidade de desembolso por unidade de produto marginal do insumo em questão, pode-se interpretar o multiplicador de Lagrange como o custo marginal de produção no estado ótimo. É claro que essa interpretação é inteiramente consistente com o que constatamos anteriormente em (12.16), isto é, que o valor ótimo do multiplicador de Lagrange mede o efeito estático comparativo da constante de restrição sobre o valor ótimo da função objetivo, isto é, $\mu^* = (\$C^*/\$Q_0)$, onde o símbolo $\$$ indica que essa é uma derivada parcial total.

A equação (12.55) pode ser escrita, alternativamente, na forma

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} \quad (12.55')$$

que você deve comparar com (12.31''). Apresentada dessa forma, a condição de primeira ordem pode ser explicada em termos de isoquantas e isocustos. Como aprendemos em (11.36), a razão Q_a/Q_b é a negativa da inclinação de uma isoquanta; ou seja, ela é uma medida da *taxa marginal de substituição técnica de a por b* ($MRTS_{ab}$). No presente modelo, o nível de produção é especificado em Q_0 ; assim, apenas uma isoquanta está envolvida, como mostra a Figura 12.8, cuja inclinação é negativa.

FIGURA 12.8



A razão P_a/P_b , por outro lado, representa a negativa da inclinação de *isocustos* (uma noção comparável à linha de orçamento na teoria do consumidor). Um isocusto, definido como o lugar geométrico das combinações de insumos que acarretam o mesmo custo total, pode ser expresso pela equação

$$C_0 = aP_a + bP_b \quad \text{ou} \quad b = \frac{C_0}{P_b} - \frac{P_a}{P_b}a$$

onde C_0 representa um número de custo (paramétrico). Quando desenhado no plano ab , como na Figura 12.8, portanto, ele resulta em uma família de retas com inclinação (negativa) $-P_a/P_b$ (e interseção com o eixo vertical C_0/P_b). A igualdade das duas razões, por conseguinte, equivale à igualdade das inclinações da isoquanta e um isocusto selecionado. Visto que somos obrigados a permanecer sobre a isoquanta dada, essa condição nos leva ao ponto de tangência E e à combinação de insumos (a^*, b^*) .

Condição de segunda ordem

Para assegurar um custo *mínimo*, é suficiente (após cumprir a condição de primeira ordem) ter um hessiano aumentado negativo, isto é, ter

$$|\overline{H}| = \begin{vmatrix} 0 & Q_a & Q_b \\ Q_a & -\mu Q_{aa} & -\mu Q_{ab} \\ Q_b & -\mu Q_{ba} & -\mu Q_{bb} \end{vmatrix} = \mu (Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2) < 0$$

Uma vez que o valor ótimo de μ (custo marginal) é positivo, isso se reduz à condição de a expressão entre parênteses ser negativa quando calculada em E .

De (11.44), lembramos que a curvatura de uma isoquanta é representada pela derivada segunda

$$\frac{d^2b}{da^2} = \frac{-1}{Q_b^3} (Q_{aa}Q_b^2 - 2Q_{ab}Q_aQ_b + Q_{bb}Q_a^2)$$

na qual aparece a mesma expressão entre parênteses. Considerando que Q_b é positiva, a satisfação da condição suficiente de segunda ordem implicaria que d^2b/da^2 é positiva – isto é, a isoquanta é estritamente convexa – no ponto de tangência. No presente contexto, a convexidade estrita da isoquanta também implicaria a satisfação da condição suficiente de segunda ordem. Pois, visto que a isoquanta tem inclinação negativa, a convexidade estrita pode significar somente uma d^2b/da^2 positiva (d^2b/da^2 de valor zero só é possível em um ponto estacionário da isoquanta), o

que, por sua vez, garantiria que $|\bar{H}| < 0$. Contudo, mais uma vez, devemos ter em mente que a condição suficiente $|\bar{H}| < 0$ (e, portanto, a convexidade estrita da isoquanta) na tangência é, por si, não necessária para a minimização de C . Especificamente, C pode ser minimizada mesmo quando a isoquanta for (não estritamente) convexa, em uma situação de vários mínimos análoga à da Figura 12.7b, com $d^2b/da^2 = 0$ e $|\bar{H}| = 0$ em cada mínimo.

Quando discutimos o modelo de maximização da utilidade (Seção 12.5), salientamos que uma função utilidade $U = U(x, y)$ suave, crescente, estritamente quase-côncava dá origem a curvas de indiferença estritamente convexas, de inclinação decrescente no plano xy . Visto que a noção de isoquantas é quase idêntica à de curvas de indiferença,[†] podemos raciocinar, por analogia, que uma função produção $Q = Q(a, b)$ suave, crescente, estritamente quase-côncava pode gerar, no plano ab , isoquantas de inclinação descendente, estritamente convexas em toda a sua extensão. Se admitirmos tal função produção, então, obviamente, a condição suficiente de segunda ordem será sempre satisfeita. Além do mais, deve ficar claro que o C^* resultante será um mínimo absoluto único restrito.

A rota de expansão

Agora vamos voltar nossa atenção aos aspectos de estática comparativa do modelo. Supondo uma razão *fixa* entre os dois preços de insumos, vamos postular aumentos sucessivos de Q_0 (elevação para isoquantas cada vez mais altas) e traçar o efeito da combinação de custo mínimo b^*/a^* . Cada deslocamento da isoquanta, é claro, resultará em um novo ponto de tangência, com um isocusto mais alto. O lugar geométrico desses pontos de tangência, conhecido como a *rota de expansão* da empresa, serve para descrever as combinações de custo mínimo requeridas para produzir níveis variados de Q_0 . Duas formas possíveis da rota de expansão são mostradas na Figura 12.9.

Se admitirmos a convexidade estrita das isoquantas (por conseguinte, a satisfação da condição de segunda ordem), a rota de expansão poderá ser diretamente deduzida da condição de primeira ordem (12.55'). Vamos ilustrar isso para a versão generalizada da função produção de Cobb-Douglas.

A condição (12.55') requer a igualdade da razão insumo-preço e da razão produto marginal. Para a função $Q = Aa^\alpha b^\beta$, isso significa que cada ponto na rota de expansão tem de satisfazer

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = \frac{A\alpha a^{\alpha-1} b^\beta}{Aa^\alpha \beta b^{\beta-1}} = \frac{\alpha b}{\beta a} \quad (12.56)$$

implicando que a razão insumo ótimo deve ser

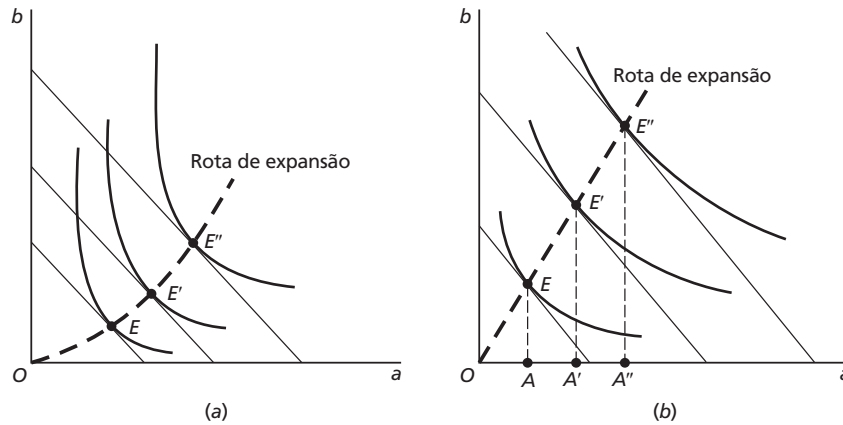
$$\frac{b^*}{a^*} = \frac{\beta P_a}{\alpha P_b} = \text{uma constante} \quad (12.57)$$

visto que α, β e os preços dos insumos são todos constantes. O resultado é que todos os pontos na rota de expansão devem mostrar a mesma razão de insumos *fixa*, isto é, a rota de expansão deve ser uma linha reta que emana do ponto de origem. Isso é ilustrado na Figura 12.9b, onde as razões de insumos nos vários pontos de tangência (AE/OA , $A'E'/OA'$ e $A''E''/OA''$) são todas iguais.

A linearidade da rota de expansão é característica da função de Cobb-Douglas generalizada, quer $\alpha + \beta = 1$ ou não, porque a dedução do resultado em (12.57) não depende da premissa $\alpha + \beta = 1$. A propósito, qualquer função produção homogênea (não necessariamente a de Cobb-Douglas) dará origem a uma rota de expansão linear para cada conjunto de preços de insumos, pela seguinte razão: se ela for homogênea de (digamos) grau r , ambas as funções produto marginal Q_a e Q_b devem ser homogêneas de grau $(r-1)$ nos insumos a e b ; assim, um aumento de j vezes em ambos os insumos produzirá uma variação de j^{r-1} vezes nos valores de *ambos*, Q_a e Q_b , o que não afetaria a razão Q_a/Q_b . Portanto, se a condição de primeira ordem $P_a/P_b = Q_a/Q_b$ for satisfeita para determinados preços de insumos por uma combinação particular de insumos (a_0, b_0) , ela também deve ser satisfeita por uma combinação (ja_0, jb_0) – exatamente como é retratado pela rota de expansão linear na Figura 12.9b.

[†] Ambas têm as características de curvas de “isovalor”. Diferem apenas na área de aplicação; curvas de indiferença são utilizadas em modelos de consumo, e isoquantas, em modelos de produção.

FIGURA 12.9



Embora *qualquer* função produção homogênea possa dar origem a uma rota de expansão linear, o grau específico de homogeneidade realmente faz uma diferença significativa na interpretação da rota de expansão. Na Figura 12.9b, desenhamos a distância OE igual à de EE' , de modo que o ponto E' envolve a duplicação da escala do ponto E . Agora, se a função produção for homogênea de grau *um*, o produto em E' deve ser duas vezes ($2^1 = 2$) o produto em E . Mas, se o grau de homogeneidade for *dois*, o produto em E' será quatro vezes ($2^2 = 4$) o produto em E . Assim, o espaçamento das isoquantas para $Q = 1, Q = 2, \dots$, será muito diferente para graus diferentes de homogeneidade.

Funções homotéticas

Já explicamos que, dado um conjunto de preços de insumos, a homogeneidade (de qualquer grau) da função produção produz uma rota de expansão linear. Mas as rotas de expansão lineares não são exclusivas de funções produção homogêneas, pois uma classe mais geral de funções, conhecida como *funções homotéticas*, também pode produzi-las.

A homoteticidade pode surgir de uma função composta na forma

$$H = h[Q(a, b)] \quad [h'(Q) \neq 0] \quad (12.58)$$

onde $Q(a, b)$ é homogênea de grau r . Embora derivada de uma função homogênea, a função $H = H(a, b)$ é, em geral, *não* homogênea nas variáveis a e b . Não obstante, as rotas de expansão de $H(a, b)$, assim como as de $Q(a, b)$, são lineares. A chave desse resultado é que, em qualquer ponto dado no plano ab , a isoquanta H compartilha a mesma inclinação da isoquanta Q :

$$\begin{aligned} \text{Inclinação da isoquanta } H &= -\frac{H_a}{H_b} = -\frac{h'(Q)Q_a}{h'(Q)Q_b} \\ &= -\frac{Q_a}{Q_b} = \text{inclinação da isoquanta } Q \end{aligned} \quad (12.59)$$

Agora a linearidade das rotas de expansão de $Q(a, b)$ implica a condição (e é implicada por ela)

$$-\frac{Q_a}{Q_b} = \text{constante para qualquer } \frac{b}{a} \text{ dada}$$

Em vista de (12.59), contudo, temos, imediatamente,

$$-\frac{H_a}{H_b} = \text{constante para qualquer } \frac{b}{a} \text{ dada} \quad (12.60)$$

também. E isso estabelece que $H(a, b)$ também produz rotas de expansão lineares.

O conceito de homoteticidade é mais geral que o de homogeneidade. Na verdade, cada função homogênea é automaticamente um membro da família homotética, mas uma função homotética pode ser uma função fora da família homogênea. O fato de uma função homogênea ser sempre homotética pode ser verificado em (12.58), onde, se deixarmos a função $H = h(Q)$ tomar a forma específica $H = Q$ – com $h'(Q) = dH/dQ = 1$ – então a função Q , sendo idêntica à função H , é obviamente homotética. O fato de que uma função homotética pode não ser homogênea será ilustrado no Exemplo 2 que virá mais adiante.

Ao definir a função homotética H , especificamos em (12.58) que $h'(Q) \neq 0$. Isso nos habilita a evitar a divisão por zero em (12.59). Conquanto a especificação $h'(Q) \neq 0$ seja o único requisito do ponto de vista matemático, considerações econômicas sugeririam a restrição mais forte $h'(Q) > 0$. Pois, se quisermos que $H(a, b)$, assim como $Q(a, b)$, sirva como uma função produção, isto é, se quisermos que H denote produção, então devemos fazer com que H_a e H_b , respectivamente, sigam a mesma direção de Q_a e Q_b na função $Q(a, b)$. Assim, $H(a, b)$ precisa se restringir a ser uma transformação monotonicamente crescente de $Q(a, b)$.

Funções de produção homotéticas (incluindo o caso especial das homogêneas) possuem a interessante propriedade de que a elasticidade (parcial) de nível ótimo de insumo em relação ao nível de produção é uniforme para todos os insumos. Para verificar isso, lembre-se de que a linearidade das rotas de expansão de funções homotéticas significa que a razão de insumos ótimos b^*/a^* não é afetada por uma variação no nível exógeno de produção H_0 . Assim, $\partial(b^*/a^*)/\partial H_0 = 0$ ou

$$\frac{1}{a^{*2}} \left(a^* \frac{\partial b^*}{\partial H_0} - b^* \frac{\partial a^*}{\partial H_0} \right) = 0 \quad [\text{regra do quociente}]$$

Multiplicando tudo por $a^{*2}H_0$ e rearrumando, obtemos, então

$$\frac{\partial a^*}{\partial H_0} \frac{H_0}{a^*} = \frac{\partial b^*}{\partial H_0} \frac{H_0}{b^*} \quad \text{ou} \quad \varepsilon_{a^*H_0} = \varepsilon_{b^*H_0}$$

que é o que afirmamos anteriormente.

EXEMPLO 1

Seja $H = Q^2$, onde $Q = Aa^\alpha b^\beta$. Visto que $Q(a, b)$ é homogênea e $h'(Q) = 2Q$ é positiva para produção positiva, $H(a, b)$ é homotética para $Q > 0$. Verificaremos que ela satisfaz (12.60). Primeiro, por substituição, temos

$$H = Q^2 = (Aa^\alpha b^\beta)^2 = A^2 a^{2\alpha} b^{2\beta}$$

Assim, a inclinação das isoquantas de H é expressa por

$$-\frac{H_a}{H_b} = -\frac{A^2 2\alpha a^{2\alpha-1} b^{2\beta}}{A^2 a^{2\alpha} 2\beta b^{2\beta-1}} = -\frac{\alpha b}{\beta a} \quad (12.61)$$

Esse resultado satisfaz (12.60) e implica rotas de expansão lineares. Comparar (12.61) com (12.56) também mostra que a função H satisfaz (12.59).

Neste exemplo, $Q(a, b)$ é homogênea de grau $(\alpha + \beta)$. Resulta que $H(a, b)$ também é homogênea, mas de grau $2(\alpha + \beta)$. Contudo, como regra, uma função homotética não é necessariamente homogênea.

EXEMPLO 2

Seja $H = e^Q$, onde $Q = Aa^\alpha b^\beta$. Visto que $Q(a, b)$ é homogênea e $h'(Q) = e^Q$ é positiva, $H(a, b)$ é homotética. Dessa função,

$$H(a, b) = \exp(Aa^\alpha b^\beta)$$

constatamos com facilidade que

$$-\frac{H_a}{H_b} = -\frac{A\alpha a^{\alpha-1}b^\beta \exp(Aa^\alpha b^\beta)}{Aa^\alpha \beta b^{\beta-1} \exp(Aa^\alpha b^\beta)} = -\frac{\alpha b}{\beta a}$$

É claro que esse resultado é idêntico a (12.61) no Exemplo 1. Dessa vez, contudo, a função homotética *não* é homogênea, porque

$$\begin{aligned} H(ja, jb) &= \exp[A(ja)^\alpha (jb)^\beta] = \exp(Aa^\alpha b^\beta j^{\alpha+\beta}) \\ &= [\exp(Aa^\alpha b^\beta)]^{j^{\alpha+\beta}} = [H(a, b)]^{j^{\alpha+\beta}} \neq j^\alpha H(a, b) \end{aligned}$$

Elasticidade de substituição

Um outro aspecto da estática comparativa tem a ver com o efeito de uma variação na razão P_a/P_b sobre a combinação de insumos de custo mínimo b^*/a^* para gerar a mesma produção dada Q_0 (isto é, enquanto permaneceremos sobre a mesma isoquanta).

Quando a razão insumo-preço (exógena) P_a/P_b aumenta, normalmente podemos esperar que a razão de insumos ótima b^*/a^* também aumente, porque o insumo b (agora relativamente mais barato) tenderá a ser substituído pelo insumo a . A *direção* da substituição é clara, mas o que dizer de sua *extensão*? A extensão da substituição de insumo pode ser medida pela seguinte expressão de elasticidade pontual, denominada *elasticidade de substituição*, denotada por σ (letra grega minúscula sigma, que representa “substituição”):

$$\sigma \equiv \frac{\text{variação relativa em } (b^*/a^*)}{\text{variação relativa em } (P_a/P_b)} = \frac{\frac{d(b^*/a^*)}{b^*/a^*}}{\frac{d(P_a/P_b)}{P_a/P_b}} = \frac{d(b^*/a^*)}{d(P_a/P_b)} \cdot \frac{P_a/P_b}{b^*/a^*} \quad (12.62)$$

O valor de σ pode estar em qualquer lugar entre 0 e ∞ ; quanto maior o σ , maior a condição de substituição entre os dois insumos. O caso limite de $\sigma = 0$ é quando os dois insumos devem ser usados segundo uma proporção fixa, como complementos um do outro. O outro caso limite, com σ infinito, é quando os dois insumos são substitutos perfeitos um do outro. Note que, se (b^*/a^*) for considerada uma função de (P_a/P_b) , então a elasticidade σ novamente será a razão entre uma função *marginal* e uma função *média*.[†]

Como ilustração, vamos calcular a elasticidade de substituição para a função produção generalizada de Cobb-Douglas. Já aprendemos anteriormente que, para esse caso, a combinação de insumos de custo mínimo é especificada por

$$\left(\frac{b^*}{a^*}\right) = \frac{\beta}{\alpha} \left(\frac{P_a}{P_b}\right) \quad [\text{de (12.57)}]$$

Essa equação está na forma $y = ax$, para a qual dy/dx (a marginal) e y/x (a média) são ambas iguais à constante a . Isto é,

[†] Há um modo alternativo de expressar σ . Visto que no ponto de tangência sempre temos

$$\frac{P_a}{P_b} = \frac{Q_a}{Q_b} = MRTS_{ab}$$

a elasticidade de substituição pode ser definida, de modo equivalente, como

$$\sigma = \frac{\text{variação relativa em } (b^*/a^*)}{\text{variação relativa em } MRTS_{ab}} = \frac{\frac{d(b^*/a^*)}{b^*/a^*}}{\frac{d(Q_a/Q_b)}{Q_a/Q_b}} = \frac{d(b^*/a^*)}{d(Q_a/Q_b)} \cdot \frac{Q_a/Q_b}{b^*/a^*} \quad (12.62')$$

$$\frac{d(b^*/a^*)}{d(P_a/P_b)} = \frac{\beta}{\alpha} \quad \text{e} \quad \frac{b^*/a^*}{P_a/P_b} = \frac{\beta}{\alpha}$$

Substituindo esses valores em (12.62), constatamos imediatamente que $\sigma = 1$; isto é, a função produção de Cobb-Douglas generalizada é caracterizada por uma elasticidade de substituição *constante e unitária*. Note que a obtenção desse resultado não depende, de modo algum, da premissa de que $\alpha + \beta = 1$. Assim, a elasticidade de substituição da função produção $Q = Aa^\alpha b^\beta$ será unitária mesmo que $\alpha + \beta \neq 1$.

Função produção CES

Mais recentemente, uma outra forma de função de produção começou a ser comumente utilizada. Essa função, embora ainda caracterizada por uma elasticidade de substituição constante (*constant elasticity of substitution* – CES), pode resultar em um σ com um valor (constante) diferente de 1.[†] A equação dessa função, conhecida como a *função produção CES* é

$$Q = A[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{-1/\rho} \quad (A > 0; 0 < \delta < 1; -1 < \rho \neq 0) \quad (12.63)$$

onde K e L representam dois fatores de produção, e A , δ e ρ (letra grega minúscula rô) são três parâmetros. O parâmetro A (*parâmetro de eficiência*) desempenha o mesmo papel que o coeficiente A na função de Cobb-Douglas; serve como um indicador do estado da tecnologia. O parâmetro δ (*parâmetro de distribuição*), assim como o α na função de Cobb-Douglas, tem a ver com as participações relativas dos fatores no produto. E o parâmetro ρ (*parâmetro de substituição*) – que não tem contraparte na função de Cobb-Douglas – é o que determina o valor da elasticidade de substituição (constante), como será demonstrado mais adiante nesta seção.

Em primeiro lugar, contudo, vamos observar que essa função é homogênea de grau um. Se substituirmos K e L por jK e jL , respectivamente, o resultado mudará de Q para

$$\begin{aligned} A[\delta(jK)^{-\rho} + (1 - \delta)(jL)^{-\rho}]^{-1/\rho} &= A\{j^{-\rho}[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]\}^{-1/\rho} \\ &= (j^{-\rho})^{-1/\rho} Q = jQ \end{aligned}$$

Por conseqüência, a função CES, como todas as funções produção linearmente homogêneas, apresenta retornos constantes de escala, qualifica-se para a aplicação do teorema de Euler e tem produtos médios e produtos marginais homogêneos de grau zero nas variáveis K e L .

Podemos observar, também, que as inclinações das isoquantas geradas pela função produção CES são sempre negativas e estritamente convexas para valores positivos de K e L . Para demonstrar isso, em primeiro lugar, vamos encontrar as expressões para os produtos marginais Q_L e Q_K . Usando a notação abreviada $[\dots]$ para representar $[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]$, temos

$$\begin{aligned} Q_L &\equiv \frac{\partial Q}{\partial L} = A \left(-\frac{1}{\rho} \right) [\dots]^{-(1/\rho)-1} (1 - \delta)(-\rho)L^{-\rho-1} \\ &= (1 - \delta) A [\dots]^{-(1+\rho)/\rho} L^{-(1+\rho)} \\ &= (1 - \delta) \frac{A^{1+\rho}}{A^\rho} [\dots]^{-(1+\rho)/\rho} L^{-(1+\rho)} \\ &= \frac{(1 - \delta)}{A^\rho} \left(\frac{Q}{L} \right)^{1+\rho} > 0 \quad [\text{por (12.63)}] \end{aligned} \quad (12.64)$$

[†] Arrow, K. J.; Chenery, H. B.; Minhas, B. S. e Solow, R. M. "Capital-Labor Substitution and Economic Efficiency". *Review of Economics and Statistics*, p. 225–250, ago. 1961.

e, de modo semelhante,

$$Q_K \equiv \frac{\partial Q}{\partial K} = \frac{\delta}{A^\rho} \left(\frac{Q}{K} \right)^{1+\rho} > 0 \quad (12.65)$$

que são definidas para valores positivos de K e L . Assim, a inclinação das isoquantas (quando representadas graficamente com K na vertical e L na horizontal) é

$$\frac{dK}{dL} = -\frac{Q_L}{Q_K} = -\frac{(1-\delta)}{\delta} \left(\frac{K}{L} \right)^{1+\rho} < 0 \quad [\text{veja (11.36)}] \quad (12.66)$$

Então, é fácil verificar que $d^2K/dL^2 > 0$ (o que deixamos para você fazer como exercício), implicando que as isoquantas são estritamente convexas para K e L positivos.

Também pode ser demonstrado que a função produção CES é estritamente quase-côncava para K e L positivos. Prosseguindo na diferenciação de (12.64) e (12.65), constatamos que as derivadas segundas da função têm os seguintes sinais:

$$Q_{LL} = \frac{\partial}{\partial L} Q_L = \frac{(1-\delta)(1+\rho)}{A^\rho} \left(\frac{Q}{L} \right)^\rho \frac{Q_L L - Q}{L^2} < 0$$

[$Q_L L - Q < 0$, pelo teorema de Euler]

$$Q_{KK} = \frac{\partial}{\partial K} Q_K = \frac{\delta(1+\rho)}{A^\rho} \left(\frac{Q}{K} \right)^\rho \frac{Q_K K - Q}{K^2} < 0$$

[$Q_K K - Q < 0$, pelo teorema de Euler]

$$Q_{KL} = Q_{LK} = \frac{(1-\delta)(1+\rho)}{A^\rho} \left(\frac{Q}{L} \right)^\rho \frac{Q_K}{L} > 0$$

Esses sinais de derivadas, válidos para K e L positivos, nos habilitam a verificar a condição suficiente para a quase-concavidade estrita (12.26). Como você pode verificar,

$$|B_1| = -Q_K^2 < 0$$

$$\text{e} \quad |B_2| = 2Q_K Q_L Q_{KL} - Q_K^2 Q_{LL} - Q_L^2 Q_{KK} > 0$$

Assim, a função CES é estritamente quase-côncava para K e L positivos.

Por fim, vamos usar os produtos marginais em (12.64) e (12.65) para achar a elasticidade de substituição da função CES. Para satisfazer a condição de combinação de custo mínimo $Q_L/Q_K = P_L/P_K$, onde P_L e P_K denotam os preços do serviço da mão-de-obra (taxa salarial) e do serviço do capital (taxa de aluguel para bens de capital), respectivamente, devemos ter

$$\frac{1-\delta}{\delta} \left(\frac{K}{L} \right)^{1+\rho} = \frac{P_L}{P_K} \quad [\text{veja (12.66)}]$$

Assim, a razão ótima de insumos é (introduzindo um símbolo abreviado c)

$$\left(\frac{K^*}{L^*} \right) = \left(\frac{\delta}{1-\delta} \right)^{1/(1+\rho)} \left(\frac{P_L}{P_K} \right)^{1/(1+\rho)} \equiv c \left(\frac{P_L}{P_K} \right)^{1/(1+\rho)} \quad (12.67)$$

Considerando (K^*/L^*) como uma função de (P_L/P_K) , constatamos que as funções marginal e média associadas são

$$\text{Função marginal} = \frac{d(K^*/L^*)}{d(P_L/P_K)} = \frac{c}{1+\rho} \left(\frac{P_L}{P_K} \right)^{1/(1+\rho)-1}$$

$$\text{Função média} = \frac{K^*/L^*}{P_L/P_K} = c \left(\frac{P_L}{P_K} \right)^{1/(1+\rho)-1}$$

Por conseguinte, a elasticidade de substituição é [†]

$$\sigma = \frac{\text{Função marginal}}{\text{Função média}} = \frac{1}{1+\rho} \quad (12.68)$$

Isso mostra que σ é uma constante cuja grandeza depende do valor do parâmetro ρ , como se segue:

$$\left. \begin{array}{l} -1 < \rho < 0 \\ \rho = 0 \\ 0 < \rho < \infty \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \sigma > 1 \\ \sigma = 1 \\ \sigma < 1 \end{array} \right.$$

Função de Cobb-Douglas como um caso especial da função CES

Nesse último resultado, o caso $\rho = 0$, que aparece no meio, leva a uma elasticidade de substituição unitária que, como sabemos, é característica da função de Cobb-Douglas. Isso sugere que a função de Cobb-Douglas (linearmente homogênea) é um caso especial da função CES (linearmente homogênea). A dificuldade é que a função CES, como dada em (12.63), é indefinida quando $\rho = 0$, porque a divisão por zero não é possível. Não obstante, podemos demonstrar que, quando $\rho \rightarrow 0$, a função CES se aproxima da função de Cobb-Douglas.

Para essa demonstração, dependeremos de uma técnica conhecida como *regra de L'Hôpital*. Essa regra tem a ver com o cálculo do limite de uma função $f(x) = \frac{m(x)}{n(x)}$ quando $x \rightarrow a$ (onde a pode ser finito ou infinito), quando o numerador $m(x)$ e o denominador $n(x)$ (1) ambos tendem a zero quando $x \rightarrow a$, resultando, assim, na expressão na forma $0/0$; ou (2) ambos tendem a $\pm\infty$ quando $x \rightarrow a$, resultando, assim, na expressão na forma de ∞/∞ (ou $\infty/-\infty$, ou $-\infty/\infty$, ou $-\infty/-\infty$). Embora o limite de $f(x)$ não possa ser calculado por essas expressões indeterminadas, ainda assim seu valor pode ser encontrado usando a fórmula

$$\lim_{x \rightarrow a} \frac{m(x)}{n(x)} = \lim_{x \rightarrow a} \frac{m'(x)}{n'(x)} \quad [\text{regra de L'Hôpital}] \quad (12.69)$$

[†] É claro que também poderíamos ter obtido o mesmo resultado calculando primeiramente os logaritmos de ambos os lados de (12.67):

$$\ln \left(\frac{K^*}{L^*} \right) = \ln c + \frac{1}{1+\rho} \ln \left(\frac{P_L}{P_K} \right)$$

e então aplicando a fórmula da elasticidade em (10.28), para obter

$$\sigma = \frac{d(\ln K^*/L^*)}{d(\ln P_L/P_K)} = \frac{1}{1+\rho}$$

EXEMPLO 3

Encontre o limite de $(1 - x^2)/(1 - x)$ quando $x \rightarrow 1$. Aqui, ambos, $m(x)$ e $n(x)$, aproximam-se de zero à medida que x se aproxima de um, exemplificando, assim, a circunstância (1). Visto que $m'(x) = -2x$ e $n'(x) = -1$, podemos escrever

$$\lim_{x \rightarrow 1} \frac{1 - x^2}{1 - x} = \lim_{x \rightarrow 1} \frac{-2x}{-1} = \lim_{x \rightarrow 1} 2x = 2$$

Essa resposta é idêntica àquela obtida por um outro método no Exemplo 2 da Seção 6.4.

EXEMPLO 4

Encontre o limite de $(2x + 5)/(x + 1)$ quando $x \rightarrow \infty$. Quando x torna-se infinito, ambos, $m(x)$ e $n(x)$, tornam-se infinitos no presente caso; assim, temos aqui um exemplo da circunstância (2). Visto que $m'(x) = 2$ e $n'(x) = 1$, podemos escrever

$$\lim_{x \rightarrow \infty} \frac{2x + 5}{x + 1} = \lim_{x \rightarrow \infty} \frac{2}{1} = 2$$

Mais uma vez, a resposta é idêntica àquela obtida por um outro método no Exemplo 3 da Seção 6.4.

Pode acontecer de a expressão do lado direito em (12.69) cair novamente no formato $0/0$ ou no formato ∞/∞ , o mesmo que a expressão do lado esquerdo. Nesse caso, podemos reaplicar a regra de L'Hôpital, isto é, podemos procurar o limite de $m''(x)/n''(x)$ quando $x \rightarrow a$, e tomar este limite como nossa resposta. Também pode acontecer que, mesmo que a função dada $f(x)$, cujo limite desejamos avaliar, não esteja originalmente na forma $m(x)/n(x)$ que cai nos formatos $0/0$ ou ∞/∞ ao se tomar o limite, uma transformação adequada fará com que $f(x)$ se preste à aplicação da regra em (12.69). Esta última possibilidade pode ser ilustrada pelo problema de achar o limite da função CES (12.63) – agora vista como uma função $Q(\rho)$ – quando $\rho \rightarrow 0$.

Tal como dada, $Q(\rho)$ não está na forma de $m(\rho)/n(\rho)$. Dividindo ambos os lados de (12.63) por A , e tomando o logaritmo natural, todavia, obtemos uma expressão naquela forma, a saber,

$$\ln \frac{Q}{A} = \frac{-\ln[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]}{\rho} \equiv \frac{m(\rho)}{n(\rho)} \quad (12.70)$$

Além disso, quando $\rho \rightarrow 0$, constatamos que $m(\rho) \rightarrow -\ln(\delta + 1 - \delta) = -\ln 1 = 0$, e $n(\rho) \rightarrow 0$ também. Assim, a regra de L'Hôpital pode ser usada para encontrar o limite de $\ln(Q/A)$. Isso feito, o limite de Q também pode ser encontrado: uma vez que $Q/A = e^{\ln(Q/A)}$, de modo que $Q = A e^{\ln(Q/A)}$, segue que

$$\lim Q = \lim A e^{\ln(Q/A)} = A e^{\lim \ln(Q/A)} \quad (12.71)$$

De (12.70), vamos, em primeiro lugar, achar $m'(\rho)$ e $n'(\rho)$, como requerido pela regra de L'Hôpital. A última é simplesmente $n'(\rho) = 1$. A primeira é

$$\begin{aligned} m'(\rho) &= \frac{-1}{[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]} \frac{d}{d\rho} [\delta K^{-\rho} + (1 - \delta)L^{-\rho}] \quad [\text{regra da cadeia}] \\ &= \frac{-[-\delta K^{-\rho} \ln K - (1 - \delta)L^{-\rho} \ln L]}{[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]} \quad [\text{por (10.21')}] \end{aligned}$$

Pela regra de L'Hôpital, temos, portanto

$$\lim_{\rho \rightarrow 0} \ln \frac{Q}{A} = \lim_{\rho \rightarrow 0} \frac{m'(\rho)}{n'(\rho)} = \frac{\delta \ln K + (1 - \delta) \ln L}{1} = \ln(K^\delta L^{1-\delta})$$

Em vista desse resultado, quando ρ é elevado à potência de $\lim_{\rho \rightarrow 0} \ln(Q/A)$, o resultado é simplesmente $K^\delta L^{1-\delta}$. Daí, por (12.71), finalmente chegamos ao resultado

$$\lim_{\rho \rightarrow 0} Q = AK^\delta L^{1-\delta}$$

que mostra que, quando $\rho \rightarrow 0$, a função CES de fato tende à função de Cobb-Douglas.

EXERCÍCIO 12.7

- Suponha que as isoquantas na Figura 12.9b são obtidas de uma função produção homogênea específica $Q = Q(a, b)$. Notando que $OE = E'E''$, quais devem ser as razões entre os níveis de produção representados pelas três isoquantas se a função Q for homogênea
(a) de grau um? (b) de grau dois?
- Para o caso da Cobb-Douglas generalizada, que tipo de curva resultará se fizermos o gráfico da razão b^*/a^* em função da razão P_a/P_b ? Esse resultado depende da premissa de que $\alpha + \beta = 1$? Leia a elasticidade de substituição no gráfico dessa curva.
- A função produção CES é caracterizada por retornos decrescentes para cada insumo para todos os níveis positivos de insumo?
- Mostre que, sobre uma isoquanta da função CES, $d^2K/dL^2 > 0$.
- (a) Para a função CES, se cada fator de produção for pago de acordo com seu produto marginal, qual é a razão entre a participação da mão-de-obra no produto e a participação do capital no produto? Um valor maior de δ significa uma participação relativa maior do capital?
(b) Para a função de Cobb-Douglas, a razão entre a participação da mão-de-obra e a participação do capital depende da razão K/L ? A mesma resposta se aplica à função CES?
- (a) A função produção CES elimina $\rho = -1$. Se $\rho = -1$, contudo, qual seria a forma geral das isoquantas para K e L positivos?
(b) σ é definido para $\rho = -1$? Qual é o limite de σ quando $\rho \rightarrow -1$?
(c) Interprete economicamente os resultados para as partes (a) e (b).
- Mostre que, escrevendo a função CES como $Q = A[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{-1/\rho}$, onde $\rho > 0$ é um novo parâmetro, podemos introduzir retornos crescentes de escala e retornos decrescentes de escala.
- Calcule os seguintes limites:

(a) $\lim_{x \rightarrow 4} \frac{x^2 - x - 12}{x - 4}$

(c) $\lim_{x \rightarrow 0} \frac{5^x - e^x}{x}$

(b) $\lim_{x \rightarrow 0} \frac{e^x - 1}{x}$

(d) $\lim_{x \rightarrow \infty} \frac{\ln x}{x}$

9. Utilizando a regra de L'Hôpital, mostre que

(a) $\lim_{x \rightarrow \infty} \frac{x^n}{e^x} = 0$

(b) $\lim_{x \rightarrow 0^+} x \ln x = 0$

(c) $\lim_{x \rightarrow 0^+} x^x = 1$

CAPÍTULO 13

Tópicos adicionais de otimização

Este capítulo trata de dois tópicos importantes. O primeiro é a programação não-linear, que amplia as técnicas de otimização restrita do Capítulo 12 permitindo a presença de *restrições de desigualdade* no problema. No Capítulo 12, as restrições deviam ser satisfeitas como igualdades estritas, isto é, elas eram sempre vinculadoras. Agora vamos considerar restrições que podem não ser vinculadoras na solução, isto é, podem ser satisfeitas como desigualdades na solução.

Na segunda parte deste capítulo, voltaremos ao reino da otimização restrita clássica para discutir alguns tópicos que deixamos de abordar nos capítulos anteriores. Entre esses tópicos estão a função objetivo indireta, o teorema do envelope e o conceito de dualidade.

13.1 Programação não-linear e condições de Kuhn-Tucker

Na história do desenvolvimento metodológico, as primeiras tentativas de lidar com restrições de desigualdade concentraram-se apenas nas lineares. Como a linearidade predominava nas restrições, bem como na função objetivo, a metodologia resultante foi muito naturalmente batizada de *programação linear*. A despeito da limitação da linearidade, entretanto, pela primeira vez era possível explicitar as variáveis de escolha especificamente como não-negativas, o que é adequado em grande parte das análises econômicas. Isso representa um avanço significativo. A programação não-linear, um desenvolvimento posterior, possibilitou manipular até mesmo restrições de desigualdade não-lineares e função objetivo não-linear. Por isso, ela ocupa um lugar muito importante na metodologia de otimização.

No problema clássico de otimização, sem nenhuma restrição explícita aos sinais das variáveis de escolha e sem nenhuma desigualdade nas restrições, a condição de primeira ordem para um extremo relativo ou local é simplesmente que as derivadas parciais primeiras da função de Lagrange (suave) em relação a todas as variáveis de escolha e os multiplicadores de Lagrange sejam iguais a zero. Em programação não-linear existe um tipo semelhante de condição de primeira ordem, conhecida como as *condições de Kuhn-Tucker*.[†] Contudo, como veremos, enquanto a condição de primeira ordem clássica é sempre necessária, não se pode conceder às condições de Kuhn-Tucker o status de condições necessárias a menos que cumpram uma certa hipótese. Por outro lado, sob certas circunstâncias específicas, as condições de Kuhn-Tucker tornam-se *condições suficientes*, ou até mesmo condições *necessárias e suficientes* também.

[†] Kuhn, H. W.; Tucker, A. W. "Nonlinear Programming". In: Neyman, J. (Ed.). *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, Califórnia: University of California Press, 1951, p. 481-492.

Uma vez que as condições de Kuhn-Tucker são o resultado analítico isolado mais importante na programação linear, é essencial que tenhamos um entendimento adequado dessas condições, bem como de suas implicações. Por questão de simplicidade na exposição, desenvolveremos essas condições em duas etapas.

Etapa 1: Efeito de restrições de não-negatividade

Considere, como a primeira etapa, um problema com restrições de não-negatividade nas variáveis de escolha, mas sem nenhuma outra restrição. Tomando, em particular, o caso de uma só variável, temos

$$\begin{array}{ll} \text{Maximize} & \pi = f(x_1) \\ \text{sujeita a} & x_1 \geq 0 \end{array} \quad (13.1)$$

onde supõe-se que a função f é diferenciável. Em vista da restrição $x_1 \geq 0$, podem surgir três situações. Primeira, se ocorrer um máximo local de π no interior da região viável sombreada na Figura 13.1, tal como o ponto A na Figura 13.1a, então temos uma *solução interior*. A condição de primeira ordem neste caso é $d\pi/dx_1 = f'(x_1) = 0$, a mesma do problema clássico. Segunda, como ilustrado pelo ponto B na Figura 13.1b, um máximo local também pode ocorrer no eixo vertical, onde $x_1 = 0$. Mesmo neste segundo caso, onde temos uma *solução de fronteira*, a condição de primeira ordem $f'(x_1) = 0$ permanece válida. Contudo, há uma terceira possibilidade, ou seja, no presente contexto, um máximo local pode ocorrer no ponto C ou no ponto D na Figura 13.1c porque, para se qualificar como um máximo local no problema (13.1), basta que o ponto candidato seja mais alto que os pontos vizinhos *dentro* da região viável. Em vista desta última possibilidade, o ponto máximo em um problema como (13.1) pode ser caracterizado não apenas pela equação $f'(x_1) = 0$, mas também pela desigualdade $f'(x_1) < 0$. Por outro lado, note que a desigualdade oposta $f'(x_1) > 0$ pode ser excluída com segurança, pois, em um ponto onde a inclinação da curva é ascendente, nunca podemos ter um máximo, mesmo que esse ponto esteja localizado no eixo vertical, tal como o ponto E na Figura 13.1a.

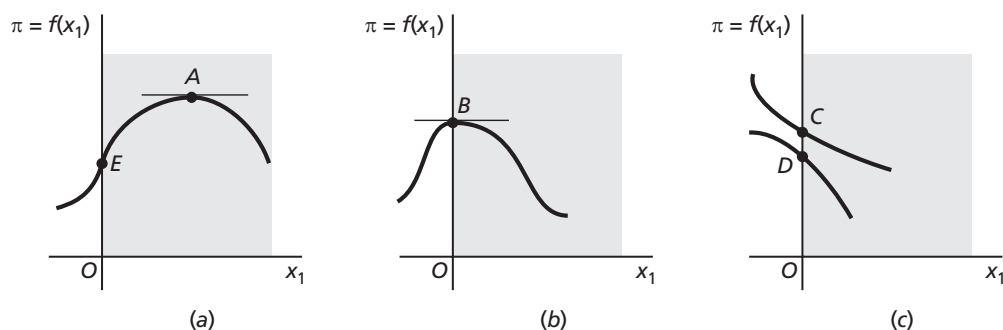
O resultado da discussão precedente é que, para que um valor de x_1 dê um máximo local de π no problema (13.1), ele deve satisfazer uma das três condições seguintes

$$f'(x_1) = 0 \quad \text{e} \quad x_1 > 0 \quad [\text{ponto } A] \quad (13.2)$$

$$f'(x_1) = 0 \quad \text{e} \quad x_1 = 0 \quad [\text{ponto } B] \quad (13.3)$$

$$f'(x_1) < 0 \quad \text{e} \quad x_1 = 0 \quad [\text{pontos } C \text{ e } D] \quad (13.4)$$

FIGURA 13.1



A propósito, essas três condições podem ser consolidadas em um único enunciado

$$f'(x_1) \leq 0 \quad x_1 \geq 0 \quad \text{e} \quad x_1 f'(x_1) = 0 \quad (13.5)$$

A primeira desigualdade em (13.5) é um resumo das informações referentes a $f'(x_1)$ enumeradas em (13.2), (13.3) e (13.4). A segunda desigualdade é um resumo semelhante para x_1 ; de fato, ela apenas reitera a restrição de não-negatividade do problema. E, quanto à terceira parte de (13.5), temos uma equação que expressa uma característica comum a (13.2), (13.3) e (13.4), a saber, das duas quantidades x_1 e $f'(x_1)$, *ao menos uma* deve assumir um valor zero, de modo que o produto das duas deve ser zero. Essa característica é denominada *folga complementar* entre x_1 e $f'(x_1)$. Tomadas em conjunto, as três partes de (13.5) constituem a condição necessária de primeira ordem para um máximo local em um problema no qual a variável de escolha deve ser não-negativa. Mas, avançando um pouco mais, também podemos considerá-las como necessárias para um máximo *global*. Isso porque um máximo global também deve ser um máximo local e, como tal, também deve satisfazer a condição necessária para um máximo local.

Quando o problema contém n variáveis de escolha:

$$\begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, \dots, x_n) \\ \text{sujeita a} & x_j \geq 0 \quad (j = 1, 2, \dots, n) \end{array} \quad (13.6)$$

A condição de primeira ordem clássica $f_1 = f_2 = \dots = f_n = 0$ deve ser modificada de maneira semelhante. Para fazer isso, podemos aplicar o mesmo tipo de raciocínio subjacente a (13.5) a cada variável de escolha x_j tomada por si. Em termos gráficos, isso equivale a ver o eixo horizontal na Figura 13.1 como representante de cada x_j por vez. Então, a modificação requerida da condição de primeira ordem se manifesta imediatamente:

$$f_j \leq 0 \quad x_j \geq 0 \quad \text{e} \quad x_j f_j = 0 \quad (j = 1, 2, \dots, n) \quad (13.7)$$

onde f_j é a derivada parcial $\partial\pi/\partial x_j$.

Etapa 2: Efeito de restrições de desigualdade

Agora que já temos um certo conhecimento básico, passamos à segunda etapa e tentamos incluir também restrições de desigualdade. Por simplicidade, em primeiro lugar vamos tratar de um problema com três variáveis de escolha ($n = 3$) e duas restrições ($m = 2$):

$$\begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, x_3) \\ \text{sujeita a} & g^1(x_1, x_2, x_3) \leq r_1 \\ & g^2(x_1, x_2, x_3) \leq r_2 \\ \text{e} & x_1, x_2, x_3 \geq 0 \end{array} \quad (13.8)$$

que, com o auxílio de duas novas variáveis, s_1 e s_2 , pode ser transformada na forma equivalente

$$\begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, x_3) \\ \text{sujeita a} & g^1(x_1, x_2, x_3) + s_1 = r_1 \\ & g^2(x_1, x_2, x_3) + s_2 = r_2 \\ \text{e} & x_1, x_2, x_3, s_1, s_2 \geq 0 \end{array} \quad (13.8')$$

Se as restrições de não-negatividade estiverem ausentes, podemos, em conformidade com a abordagem clássica, formar a função de Lagrange:

$$Z' = f(x_1, x_2, x_3) + \lambda_1[r_1 - g^1(x_1, x_2, x_3) - s_1] + \lambda_2[r_2 - g^2(x_1, x_2, x_3) - s_2] \quad (13.9)$$

e escrever a condição de primeira ordem como

$$\frac{\partial Z'}{\partial x_1} = \frac{\partial Z'}{\partial x_2} = \frac{\partial Z'}{\partial x_3} = \frac{\partial Z'}{\partial s_1} = \frac{\partial Z'}{\partial s_2} = \frac{\partial Z'}{\partial \lambda_1} = \frac{\partial Z'}{\partial \lambda_2} = 0$$

Mas, visto que as variáveis x_j e s_i têm de ser obrigatoriamente não-negativas, a condição de primeira ordem para essas variáveis deve ser modificada de acordo com (13.7). Por consequência, obtemos o seguinte conjunto de equações:

$$\begin{aligned} \frac{\partial Z'}{\partial x_j} \leq 0 & \quad x_j \geq 0 & \quad e & \quad x_j \frac{\partial Z'}{\partial x_j} = 0 \\ \frac{\partial Z'}{\partial s_i} \leq 0 & \quad s_i \geq 0 & \quad e & \quad s_i \frac{\partial Z'}{\partial s_i} = 0 \\ \frac{\partial Z'}{\partial \lambda_i} = 0 & & & \quad \left(\begin{array}{l} i = 1, 2 \\ j = 1, 2, 3 \end{array} \right) \end{aligned} \quad (13.10)$$

Note que ainda é preciso estabelecer que as derivadas $\partial Z' / \partial \lambda_i$ sejam estritamente iguais a zero. (Por quê?)

Cada linha de (13.10) está relacionada a um tipo diferente de variável. Mas podemos consolidar as duas últimas linhas e, no processo, eliminar a nova variável s_i da condição de primeira ordem. Considerando que $\partial Z' / \partial s_i = -\lambda_i$, a segunda linha de (13.10) nos diz que devemos ter $-\lambda_i \leq 0$, $s_i \geq 0$, e $-s_i \lambda_i = 0$ ou, de modo equivalente,

$$s_i \geq 0 \quad \lambda_i \geq 0 \quad e \quad s_i \lambda_i = 0 \quad (13.11)$$

Mas a terceira linha – um outro modo de enunciar as restrições em (13.8') – significa que $s_i = r_i - g^i(x_1, x_2, x_3)$. Portanto, substituindo a última em (13.11), podemos combinar a segunda e a terceira linhas de (13.10) em

$$r_i - g^i(x_1, x_2, x_3) \geq 0 \quad \lambda_i \geq 0 \quad e \quad \lambda_i [r_i - g^i(x_1, x_2, x_3)] = 0$$

Isso nos permite expressar a condição de primeira ordem (13.10) em uma forma equivalente *sem* as novas variáveis. Usando o símbolo g_j^i para denotar $\partial g^i / \partial x_j$, escrevemos agora

$$\frac{\partial Z'}{\partial x_j} = f_j - (\lambda_1 g_j^1 + \lambda_2 g_j^2) \leq 0 \quad x_j \geq 0 \quad e \quad x_j \frac{\partial Z'}{\partial x_j} = 0 \quad (13.12)$$

$$r_i - g^i(x_1, x_2, x_3) \geq 0 \quad \lambda_i \geq 0 \quad e \quad \lambda_i [r_i - g^i(x_1, x_2, x_3)] = 0$$

São essas, então, as condições de Kuhn-Tucker para o problema (13.8) ou, mais exatamente, uma versão das condições de Kuhn-Tucker expressa em termos da função de Lagrange Z' em (13.9).

Agora que conhecemos os resultados, entretanto, é possível obter o mesmo conjunto de condições de modo mais direto usando uma função de Lagrange diferente. Dado o problema

(13.9), vamos ignorar as restrições de não-negatividade, bem como os sinais de desigualdade nas restrições e escrever o tipo estritamente clássico da função de Lagrange Z :

$$Z = f(x_1, x_2, x_3) + \lambda_1[r_1 - g^1(x_1, x_2, x_3)] + \lambda_2[r_2 - g^2(x_1, x_2, x_3)] \quad (13.13)$$

Então, vamos fazer o seguinte: (1) estabelecer que as derivadas parciais $\partial Z/\partial x_j \leq 0$, mas $\partial Z/\partial \lambda_i \geq 0$; (2) impor restrições de não-negatividade a x_j e λ_i ; e (3) exigir que prevaleça a folga complementar entre cada variável e a derivada parcial de Z em relação àquela variável, isto é, exigir que seu produto seja nulo. Visto que os resultados dessas etapas, a saber,

$$\begin{aligned} \frac{\partial Z}{\partial x_j} = f_j - (\lambda_1 g_j^1 + \lambda_2 g_j^2) &\leq 0 & x_j &\geq 0 & \text{e} & x_j \frac{\partial Z}{\partial x_j} &= 0 \\ \frac{\partial Z}{\partial \lambda_i} = r_i - g^i(x_1, x_2, x_3) &\geq 0 & \lambda_i &\geq 0 & \text{e} & \lambda_i \frac{\partial Z}{\partial \lambda_i} &= 0 \end{aligned} \quad (13.14)$$

são idênticos aos de (13.12), as condições de Kuhn-Tucker também podem ser expressas em termos da função de Lagrange Z (em comparação com Z'). Note que, trocando de Z' para Z , podemos não somente chegar às condições de Kuhn-Tucker um modo mais direto, mas também identificar a expressão $r_i - g^i(x_1, x_2, x_3)$ – à qual não foi dado um nome em (13.12) – como a derivada parcial $\partial Z/\partial \lambda_i$. Por conseguinte, na discussão subsequente usaremos somente a versão (13.14) das condições de Kuhn-Tucker, com base na função de Lagrange Z .

EXEMPLO 1

Se formularmos o problema familiar de maximização de utilidade segundo o molde da programação não-linear, podemos ter um problema com a restrição de desigualdade, como se segue:

$$\begin{aligned} &\text{Maximize} && U = U(x, y) \\ &\text{sujeita a} && P_x x + P_y y \leq B \\ &\text{e} && x, y \geq 0 \end{aligned}$$

Note que, com a restrição de desigualdade, o consumidor já não é mais obrigado a gastar toda a quantia B .

Contudo, para acrescentar um novo dado ao problema, vamos supor que tenha sido imposto à mercadoria x um racionamento igual a X_0 . Assim, o consumidor enfrentaria uma segunda restrição e o problema mudaria para

$$\begin{aligned} &\text{Maximize} && U = U(x, y) \\ &\text{sujeita a} && P_x x + P_y y \leq B \\ &&& x \leq X_0 \\ &\text{e} && x, y \geq 0 \end{aligned}$$

A função de Lagrange é

$$Z = U(x, y) + \lambda_1(B - P_x x - P_y y) + \lambda_2(X_0 - x)$$

e as condições de Kuhn-Tucker são

$$\begin{aligned} Z_x = U_x - P_x \lambda_1 - \lambda_2 &\leq 0 & x &\geq 0 & \text{e} & x Z_x &= 0 \\ Z_y = U_y - P_y \lambda_1 &\leq 0 & y &\geq 0 & \text{e} & y Z_y &= 0 \\ Z_{\lambda_1} = B - P_x y - P_y y &\geq 0 & \lambda_1 &\geq 0 & \text{e} & \lambda_1 Z_{\lambda_1} &= 0 \\ Z_{\lambda_2} = X_0 - x &\geq 0 & \lambda_2 &\geq 0 & \text{e} & \lambda_2 Z_{\lambda_2} &= 0 \end{aligned}$$

É útil examinar as implicações da terceira coluna das condições de Kuhn-Tucker. A condição $\lambda_1 Z_{\lambda_1} = 0$, em particular, requer que

$$\lambda_1(B - P_x x - P_y y) = 0$$

Portanto, devemos ter ou

$$\lambda_1 = 0 \quad \text{ou} \quad B - P_x x - P_y y = 0$$

Se interpretarmos λ_1 como a utilidade marginal da verba orçamentária (renda) e se a restrição orçamentária for não-vinculadora (satisfeita como uma desigualdade na solução, com sobra de dinheiro), a utilidade marginal de B deve ser zero ($\lambda_1 = 0$).

De maneira semelhante, a condição $\lambda_2 Z_{\lambda_2} = 0$ requer que

$$\lambda_2 = 0 \quad \text{ou} \quad X_0 - x = 0$$

Visto que λ_2 pode ser interpretada como a utilidade marginal do abrandamento da restrição, vemos que, se a restrição de racionamento for não-vinculadora, a utilidade marginal do abrandamento da restrição deve ser zero ($\lambda_2 = 0$).

Essa característica, conhecida como folga complementar, desempenha um papel essencial na busca de uma solução, o que ilustraremos a seguir com um exemplo numérico:

Maximize	$U = xy$
sujeita a	$x + y \leq 100$
	$x \leq 40$
e	$x, y \geq 0$

Função de Langrange

$$Z = xy + \lambda_1(100 - x - y) + \lambda_2(40 - x)$$

e as condições de Kuhn-Tucker se tornam

$Z_x = y - \lambda_1 - \lambda_2 \leq 0$	$x \geq 0$	e	$xZ_x = 0$
$Z_y = x - \lambda_1 \leq 0$	$y \geq 0$	e	$yZ_y = 0$
$Z_{\lambda_1} = 100 - x - y \geq 0$	$\lambda_1 \geq 0$	e	$\lambda_1 Z_{\lambda_1} = 0$
$Z_{\lambda_2} = 40 - x \geq 0$	$\lambda_2 \geq 0$	e	$\lambda_2 Z_{\lambda_2} = 0$

A abordagem típica para resolver um problema de programação não-linear é a de tentativa e erro. Por exemplo, podemos começar experimentando um valor zero para uma variável de escolha. Atribuir um valor zero a uma variável sempre simplifica as condições marginais por causar a exclusão de certos termos. Então, se for possível achar valores não-negativos adequados para multiplicadores de Lagrange que satisfaçam todas as desigualdades marginais, a solução zero será ótima. Se, por outro lado, a solução zero violar algumas das desigualdades, então devemos deixar que uma ou mais variáveis de escolha sejam positivas. Para toda variável de escolha positiva, podemos, por meio da folga complementar, converter uma condição de desigualdade marginal fraca em uma igualdade restrita. Resolvida adequadamente, tal igualdade nos levará ou a uma solução ou a uma contradição que nos obrigaria a tentar alguma outra coisa. Se existir uma solução, essas tentativas paulatinamente nos habilitarão a descobri-la. Podemos também começar supondo que uma das restrições é não-vinculadora. Então o multiplicador de Lagrange relacionado será zero por folga complementar e, assim, teremos eliminado uma variável. Se essa hipótese levar a uma contradição, então devemos tratar a restrição citada como uma igualdade estrita e prosseguir nessa base.

Para o presente exemplo, não faz sentido experimentar $x = 0$ ou $y = 0$, pois teríamos $U = xy = 0$. Portanto, vamos supor que ambos, x e y , são diferentes de zero e deduzir que $Z_x = Z_y = 0$ pela folga complementar, o que significa que

$$y - \lambda_1 - \lambda_2 = x - \lambda_1 (= 0)$$

de modo que

$$y - \lambda_2 = x.$$

Agora, suponha que a restrição de racionamento seja não-vinculadora na solução, o que implica que $\lambda_2 = 0$. Então temos $x = y$, e o orçamento dado $B = 100$ resulta na solução experimental $x = y = 50$. Mas essa solução viola a restrição de racionamento $x \leq 40$. Por conseguinte, temos de adotar a hipótese alternativa de que a restrição de racionamento é vinculadora com $x^* = 40$. Então, a restrição orçamentária permite que o consumidor tenha $y^* = 60$. Além disso, visto que a folga complementar determina que $Z_x = Z_y = 0$, podemos calcular, de imediato, que $\lambda_1^* = 40$ e $\lambda_2^* = 20$.

Interpretação das condições de Kuhn-Tucker

Partes das condições de Kuhn-Tucker (13.14) são meras reafirmações de certos aspectos do problema em questão. Assim, as condições $x_j \geq 0$ apenas reproduzem as restrições de não-negatividade, e as condições $\partial Z / \partial \lambda_i \geq 0$ apenas reiteram as restrições. Contudo, incluir essas expressões em (13.14) tem a importante vantagem de revelar mais claramente a notável simetria entre os dois tipos de variáveis: x_j (variável de escolha) e λ_i (multiplicadores de Lagrange). A cada variável em cada categoria corresponde uma condição marginal – $\partial Z / \partial x_j \leq 0$ ou $\partial Z / \partial \lambda_i \geq 0$ – a ser satisfeita pela solução ótima. Cada uma das variáveis também deve ser não-negativa. E, por fim, cada variável é caracterizada por folga complementar em relação a uma derivada parcial particular da função de Lagrange Z . Isso significa que, para cada x_j , devemos achar na solução ótima que *ou* a condição marginal é válida como uma igualdade, como no contexto clássico *ou* a variável de escolha em questão deve assumir um valor zero *ou* ambas. Analogamente, para cada λ_i , devemos encontrar na solução ótima que *ou* a condição marginal é válida como uma igualdade – significando que a *i*-ésima restrição é satisfeita com exatidão – *ou* o multiplicador de Lagrange se anula *ou* ambos.

Uma interpretação ainda mais explícita é possível quando examinamos a expressão expandida para $\partial Z / \partial x_j$ e $\partial Z / \partial \lambda_i$ em (13.14). Suponha que estamos tratando do conhecido problema da produção. Então temos

$f_j \equiv$ lucro marginal bruto do *j*-ésimo produto

$\lambda_i \equiv$ preço sombra do *i*-ésimo recurso (o custo de oportunidade de usar uma unidade do *i*-ésimo recurso)

$g_j^i \equiv$ quantidade do *i*-ésimo recurso exaurido na produção da unidade marginal do *j*-ésimo produto

$\lambda_i g_j^i \equiv$ custo marginal imputado do *i*-ésimo recurso incorrido na produção de uma unidade do *j*-ésimo produto

$\sum_i \lambda_i g_j^i \equiv$ custo marginal imputado agregado do *j*-ésimo produto

Assim, a condição marginal

$$\frac{\partial Z}{\partial x_j} = f_j - \sum_i \lambda_i g_j^i \leq 0$$

requer que o lucro marginal bruto do *j*-ésimo produto não seja maior que seu custo marginal imputado agregado; isto é, não é permitida nenhuma *sub*imputação. A condição de folga complementar então significa que, se a solução ótima exigir a produção ativa do *j*-ésimo produto ($x_j^* > 0$), o lucro marginal bruto deve ser exatamente igual ao custo marginal imputado agregado ($\partial Z / \partial x_j^* = 0$), tal como seria a situação no problema clássico de otimização. Se, por outro lado, o lucro marginal bruto não alcançar otimamente o custo imputado agregado ($\partial Z / \partial x_j^* < 0$), acarretando *excesso* de imputação, esse produto não deve ser produzido ($x_j^* = 0$).[†] Esta última situação é algo que nunca pode ocorrer no contexto clássico pois, se o lucro marginal bruto for menor que o custo marginal imputado, então a produção segundo essa estrutura seria reduzida até o nível em que a condição marginal é satisfeita como uma igualdade. O que faz com que a situação de $\partial Z /$

[†] Lembre-se de que, dada a equação $ab = 0$, onde a e b são números reais, podemos inferir, com legitimidade, que $a \neq 0$ implica $b = 0$, mas não é verdade que $a = 0$ implica $b \neq 0$, já que $b = 0$ também é consistente com $a = 0$.

$\partial x_j^* < 0$ se qualifique como uma situação ótima aqui é a especificação explícita da não negatividade na atual estrutura. Então, o máximo que podemos fazer, no sentido de reduzir a da produção, é baixá-la até o nível $x_j^* = 0$ e, se ainda encontrarmos $\partial Z/\partial x_j^* < 0$ na produção zero, é aí que paramos, de qualquer modo.

Quanto às condições restantes, relacionadas às variáveis λ_i , seus significados são ainda mais fáceis de perceber. Antes de mais nada, a condição marginal $\partial Z/\partial \lambda_i \geq 0$ requer apenas que a empresa fique dentro da limitação de capacidade de cada recurso na unidade industrial. A condição de folga complementar então estipula que, se o i -ésimo recurso não for totalmente utilizado na solução ótima ($\partial Z/\partial \lambda_i^* > 0$), o preço sombra daquele recurso – que nunca se permite que seja negativo – deve ser estabelecido como igual a zero ($\lambda_i^* = 0$). Por outro lado, se um recurso tiver um preço sombra positivo na solução ótima ($\lambda_i^* > 0$), então ele é, forçosamente um recurso totalmente utilizado ($\partial Z/\partial \lambda_i^* = 0$).

É claro que também é possível considerar o valor do multiplicador de Lagrange λ_i^* como uma medida da reação do valor ótimo da função objetivo a um ligeiro abrandamento da i -ésima restrição. Sob essa perspectiva, folga complementar significaria que, se a i -ésima restrição for otimamente não-vinculadora ($\partial Z/\partial \lambda_i^* > 0$), então abrandar essa restrição particular não afetará o valor ótimo do lucro bruto ($\lambda_i^* = 0$) – exatamente como afrouxar um cinto que não está apertando nossa cintura não produzirá nenhum conforto adicional. Se, por outro lado, um ligeiro abrandamento da i -ésima restrição (aumentar a dotação do i -ésimo recurso) não aumentar o lucro bruto ($\lambda_i^* > 0$), então aquela restrição de recurso deve ser, de fato, vinculadora na solução ótima ($\partial Z/\partial \lambda_i^* = 0$).

O caso de n variáveis, m restrições

A discussão precedente pode ser generalizada diretamente para o caso em que há n variáveis de escolha e m restrições. A função de Lagrange Z aparecerá na forma mais geral

$$Z = f(x_1, x_2, \dots, x_n) + \sum_{i=1}^m \lambda_i [r_i - g^i(x_1, x_2, \dots, x_n)] \quad (13.15)$$

E as condições de Kuhn-Tucker serão, simplesmente,

$$\begin{aligned} \frac{\partial Z}{\partial x_j} \leq 0 \quad x_j \geq 0 \quad \text{e} \quad x_j \frac{\partial Z}{\partial x_j} = 0 \quad [\text{maximização}] \\ \frac{\partial Z}{\partial \lambda_i} \geq 0 \quad \lambda_i \geq 0 \quad \text{e} \quad \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \quad \begin{pmatrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{pmatrix} \end{aligned} \quad (13.16)$$

Aqui, para evitar uma aparência confusa, não escrevemos as expressões expandidas para as derivadas parciais $\partial Z/\partial x_j$ e $\partial Z/\partial \lambda_i$. Mas aconselhamos que você as escreva para ter uma visão mais detalhada das condições de Kuhn-Tucker, semelhante à que foi dada em (13.14). Note que, à parte a mudança na dimensão do problema, as condições de Kuhn-Tucker permanecem exatamente as mesmas. A interpretação dessas condições, naturalmente, também deve permanecer a mesma.

E se o problema for de *minimização*? Um modo de lidar com ele é convertê-lo em um problema de maximização e então aplicar (13.6). Minimizar C equivale a maximizar $-C$, portanto, tal conversão é sempre viável. Mas é claro que também temos de inverter as desigualdades de restrição multiplicando cada restrição por -1 . Todavia, em vez de executar o processo de conversão, podemos – mais uma vez usando a função de Lagrange Z tal como definida em (13.15) – aplicar diretamente a versão de minimização das condições de Kuhn-Tucker como se segue:

$$\frac{\partial Z}{\partial x_j} \geq 0 \quad x_j \geq 0 \quad \text{e} \quad x_j \frac{\partial Z}{\partial x_j} = 0 \quad [\text{minimização}]$$

$$\frac{\partial Z}{\partial \lambda_i} \leq 0 \quad \lambda_i \geq 0 \quad \text{e} \quad \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \quad \begin{pmatrix} i=1, 2, \dots, m \\ j=1, 2, \dots, n \end{pmatrix} \quad (13.17)$$

Compare essas expressões com (13.16).

Lendo (13.16) e (13.17) no sentido horizontal (sentido da *linha*), vemos que as condições de Kuhn-Tucker para problemas de maximização, bem como de minimização, consistem em um conjunto de condições relativas às variáveis de escolha x_j (primeira linha) e em um outro conjunto relativo aos multiplicadores de Lagrange λ_i (segunda linha). Lendo-as na vertical (no sentido da *coluna*), por outro lado, notamos que, para cada x_j e λ_i há uma condição marginal (primeira coluna), uma restrição não-negativa (segunda coluna) e uma condição de folga complementar (terceira coluna). Em qualquer problema dado, as condições marginais pertinentes às variáveis de escolha, como um grupo, sempre satisfazem desigualdades com sentido oposto às das satisfeitas pelas condições marginais para os multiplicadores de Lagrange.

Sujeitas à hipótese que será explicada na Seção 13.2, as condições de Kuhn-Tucker para máximo (13.16) e mínimo (13.17) são condições necessárias para um máximo local e um mínimo local, respectivamente. Mas, visto que um máximo (mínimo) global também tem de ser um máximo (mínimo) local, as condições de Kuhn-Tucker também podem ser tomadas como condições necessárias para um máximo (mínimo) global, sujeitas à mesma hipótese.

EXEMPLO 2

Vamos aplicar as condições de Kuhn-Tucker para resolver um problema de minimização:

$$\begin{array}{ll} \text{Minimize} & C = (x_1 - 4)^2 + (x_2 - 4)^2 \\ \text{sujeita a} & 2x_1 + 3x_2 \geq 6 \\ & -3x_1 - 2x_2 \geq -12 \\ \text{e} & x_1, x_2 \geq 0 \end{array}$$

A função de Lagrange para este problema é

$$Z = (x_1 - 4)^2 + (x_2 - 4)^2 + \lambda_1(6 - 2x_1 - 3x_2) + \lambda_2(-12 + 3x_1 + 2x_2)$$

Visto que o problema é de minimização, as condições apropriadas são (13.17), as quais incluem as quatro condições marginais

$$\begin{aligned} \frac{\partial Z}{\partial x_1} &= 2(x_1 - 4) - 2\lambda_1 + 3\lambda_2 \geq 0 \\ \frac{\partial Z}{\partial x_2} &= 2(x_2 - 4) - 3\lambda_1 + 2\lambda_2 \geq 0 \\ \frac{\partial Z}{\partial \lambda_1} &= 6 - 2x_1 - 3x_2 \leq 0 \\ \frac{\partial Z}{\partial \lambda_2} &= -12 + 3x_1 + 2x_2 \leq 0 \end{aligned} \quad (13.18)$$

mais as condições de não negatividade e de folga complementar.

Para achar uma solução, mais uma vez usamos a abordagem de tentativa e erro, tendo sempre em mente que as primeiras tentativas podem nos levar a um beco sem saída. Suponha que experimentamos, em primeiro lugar, $\lambda_1 > 0$ e $\lambda_2 > 0$ e verificamos se podemos achar os valores correspondentes de x_1 e x_2 que satisfaçam ambas as restrições. Com multiplicadores de Lagrange positivos, devemos ter $\partial Z / \partial \lambda_1 = \partial Z / \partial \lambda_2 = 0$. Assim, das duas últimas linhas de (13.18), podemos escrever

$$2x_1 + 3x_2 = 6 \quad \text{e} \quad 3x_1 + 2x_2 = 12$$

Essas duas equações resultam na solução experimental $x_1 = 4\frac{4}{5}$ e $x_2 = -1\frac{1}{5}$, que viola a restrição de não negatividade de x_2 .

Em seguida, vamos experimentar $x_1 > 0$ e $x_2 > 0$, o que implicaria $\partial Z/\partial x_1 = \partial Z/\partial x_2 = 0$ por folga complementar. Então, das duas primeiras linhas de (13.18), podemos escrever

$$2(x_1 - 4) - 2\lambda_1 + 3\lambda_2 = 0 \quad \text{e} \quad 2(x_2 - 4) - 3\lambda_1 + 2\lambda_2 = 0 \quad (13.19)$$

Multiplicando a primeira equação por 2, a segunda equação por 3 e então subtraindo a última da primeira, podemos eliminar λ_2 e obter o resultado

$$4x_1 - 6x_2 + 5\lambda_1 + 8 = 0$$

Supondo ainda, que $\lambda_1 = 0$, podemos obter a seguinte relação entre x_1 e x_2 :

$$x_1 - \frac{3}{2}x_2 = -2 \quad (13.20)$$

Para resolver para as duas variáveis, contudo, precisamos de uma outra relação entre x_1 e x_2 . Com essa finalidade, vamos supor que $\lambda_2 \neq 0$, de modo que $\partial Z/\partial \lambda_2 = 0$. Então, das últimas duas linhas de (13.18), podemos escrever (após rearranjar)

$$3x_1 + 2x_2 = 12 \quad (13.21)$$

Juntas, (13.20) e (13.21) resultam em uma outra solução experimental

$$x_1 = \frac{28}{13} \left(= 2\frac{2}{13} \right) > 0 \quad x_2 = \frac{36}{13} \left(= 2\frac{10}{13} \right) > 0$$

Substituindo esses valores em (13.19) e resolvendo para os multiplicadores de Lagrange, obtemos

$$\lambda_1 = 0 \quad \lambda_2 = \frac{16}{13} \left(= 1\frac{3}{13} \right) > 0$$

Visto que os valores de solução para as quatro variáveis são todos não-negativos e satisfazem ambas as restrições, eles são aceitáveis como a solução final.

EXERCÍCIO 13.1

- Desenhe um conjunto de diagramas semelhante ao da Figura 13.1 para o caso de minimização e deduza um conjunto de condições necessárias para um mínimo local correspondente a (13.2), (13.3) e (13.4). Então sintetize essas condições em um único enunciado semelhante a (13.5).

- (a) Mostre que, em (13.16), em vez de escrever

$$\lambda_i \frac{\partial Z}{\partial \lambda_i} = 0 \quad (i = 1, \dots, m)$$

como um conjunto de m condições separadas, basta escrever uma única equação

$$\text{na forma de } \sum_{i=1}^m \lambda_i \frac{\partial Z}{\partial \lambda_i} = 0$$

- (b) Podemos fazer o mesmo para o seguinte conjunto de condições?

$$x_j \frac{\partial Z}{\partial x_j} = 0 \quad (j = 1, \dots, n)$$

3. Com base no raciocínio usado no Problema 2, qual conjunto (ou conjuntos) de condições em (13.17) pode ser sintetizado em uma única equação?
4. Suponha que o problema é

$$\begin{aligned} &\text{Minimize} && C = f(x_1, x_2, \dots, x_n) \\ &\text{sujeita a} && g^i(x_1, x_2, \dots, x_n) \geq r_i \\ &\text{e} && x_j \geq 0 \quad \begin{pmatrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{pmatrix} \end{aligned}$$
 Escreva a função de Lagrange, tome as derivadas $\partial Z / \partial x_j$ e $\partial Z / \partial \lambda_i$ e escreva a versão expandida das condições de Kuhn-Tucker para mínimo (13.17).
5. Converta o problema de minimização do Problema 4 em um problema de maximização, formule a função de Lagrange, tome as derivadas com relação a x_j e λ_i e aplique as condições de Kuhn-Tucker para máximo (13.16). Os resultados são consistentes com os obtidos no Problema 4?

13.2 A qualificação da restrição

As condições de Kuhn-Tucker são condições necessárias *somente se* uma hipótese específica for satisfeita. Essa hipótese, denominada *qualificação da restrição*, impõe uma certa restrição às funções de restrição de um problema de programação não-linear, com o propósito específico de excluir certas irregularidades na fronteira do conjunto viável, que invalidariam as condições de Kuhn-Tucker caso a solução ótima ocorresse ali.

Irregularidades nos pontos de fronteira

Em primeiro lugar, vamos ilustrar a natureza dessas irregularidades por meio de alguns exemplos concretos.

EXEMPLO 1

$$\begin{aligned} &\text{Maximize} && \pi = x_1 \\ &\text{sujeita a} && x_2 - (1 - x_1)^3 \leq 0 \\ &\text{e} && x_1, x_2 \geq 0 \end{aligned}$$

Como mostra a Figura 13.2, a região viável é o conjunto de pontos que estão no primeiro quadrante, sobre ou abaixo da curva $x_2 = (1 - x_1)^3$. Visto que a função objetivo nos ordena maximizar x_1 , a solução ótima é o ponto $(1, 0)$. Mas a solução não satisfaz as condições de Kuhn-Tucker para máximo. Para verificar isso, em primeiro lugar escrevemos a função de Lagrange

$$Z = x_1 + \lambda_1[-x_2 + (1 - x_1)^3]$$

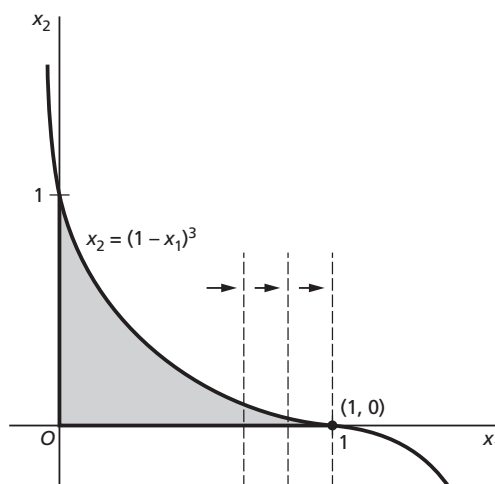
Então devemos ter, como a primeira condição marginal

$$\frac{\partial Z}{\partial x_1} = 1 - 3\lambda_1(1 - x_1)^2 \leq 0$$

De fato, visto que $x_1^* = 1$ é positivo, a folga complementar requer que essa derivada se anule quando calculada no ponto $(1, 0)$. Contudo, acontece que o valor real que obtemos é $\partial Z / \partial x_1^* = 1$, o que viola a condição marginal dada.

A razão para essa anomalia é que, neste exemplo, a solução ótima $(1, 0)$ ocorre em uma *cúspide* voltada para fora, o que constitui um tipo de irregularidade que pode invalidar as condições de Kuhn-Tucker em uma solução ótima de fronteira. Uma *cúspide* é um bico formado quando uma curva muda repentinamente de sentido, de modo que a inclinação da curva de um lado do ponto é a mesma que a inclinação da curva do outro lado do ponto. Aqui, a fronteira da região viável pri-

FIGURA 13.2



meiramente segue a curva de restrição, mas, quando alcança o ponto $(1, 0)$, ela se volta abruptamente para a esquerda (ou na direção oeste) e dali em diante acompanha a direção do eixo horizontal. Visto que a inclinação do lado curvo, bem como a do lado horizontal da fronteira é zero no ponto $(1, 0)$, este ponto é uma cuspide.

Cúspides são as mais citadas como culpadas pela falha das condições de Kuhn-Tucker, mas a verdade é que a presença de uma cuspide não é nem necessária nem suficiente para fazer com que essas condições falhem em uma solução ótima. Os Exemplos 2 e 3 confirmarão isso.

EXEMPLO 2

Vamos adicionar uma nova restrição ao problema do Exemplo 1,

$$2x_1 + x_2 \leq 2$$

cujas fronteira, $x_2 = 2 - 2x_1$, é representada graficamente por uma reta de inclinação -2 que contém o ponto ótimo na Figura 13.2. A região viável permanece claramente a mesma de antes, assim como a solução ótima na cuspide. Mas, se escrevermos a nova função de Lagrange

$$Z = x_1 + \lambda_1[-x_2 + (1 - x_1)^3] + \lambda_2[2 - 2x_1 - x_2]$$

e as condições marginais

$$\frac{\partial Z}{\partial x_1} = 1 - 3\lambda_1(1 - x_1)^2 - 2\lambda_2 \leq 0$$

$$\frac{\partial Z}{\partial x_2} = -\lambda_1 - \lambda_2 \leq 0$$

$$\frac{\partial Z}{\partial \lambda_1} = -x_2 + (1 - x_1)^3 \geq 0$$

$$\frac{\partial Z}{\partial \lambda_2} = 2 - 2x_1 - x_2 \geq 0$$

Resulta que os valores $x_1^* = 1$, $x_2^* = 0$, $\lambda_1^* = 1$ e $\lambda_2^* = \frac{1}{2}$ satisfazem essas quatro desigualdades, bem como as condições de não-negatividade e de folga complementar. Aliás, pode-se atribuir qualquer valor não-negativo a λ_1^* (não apenas 1) e, ainda assim, todas as condições podem ser satisfeitas – o que mostra que o valor ótimo de um multiplicador de Lagrange não é necessariamente único. Porém, o mais importante é que esse exemplo mostra que as condições de Kuhn-Tucker podem permanecer válidas a despeito da cuspide.

EXEMPLO 3 A região viável do problema

$$\begin{aligned}
 &\text{Maximize} && \pi = x_2 - x_1^2 \\
 &\text{sujeita a} && -(10 - x_1^2 - x_2)^3 \leq 0 \\
 &&& -x_1 \leq -2 \\
 &\text{e} && x_1, x_2 \geq 0
 \end{aligned}$$

como mostra a Figura 13.3, não contém nenhuma cúspide em nenhum lugar. Mesmo assim, as condições de Kuhn-Tucker não são válidas na solução ótima (2, 6). Pois, com a função de Lagrange

$$Z = x_2 - x_1^2 + \lambda_1 (10 - x_1^2 - x_2)^2 + \lambda_2 (-2 + x_1)$$

a segunda condição marginal exigiria que

$$\frac{\partial Z}{\partial x_2} = 1 - 3\lambda_1 (10 - x_1^2 - x_2)^2 \leq 0$$

Realmente, posto que x_2^* é positivo, essa derivada deve se anular quando calculada no ponto (2, 6). Mas, na verdade, obtemos $\partial Z / \partial x_2 = 1$, independentemente do valor atribuído a λ_1 . Assim, as condições de Kuhn-Tucker podem falhar mesmo na ausência de uma cúspide – ou melhor, até mesmo quando a região viável for um conjunto convexo, como na Figura 13.3. A razão fundamental por que cúspides não são nem necessárias nem suficientes para a falha das condições de Kuhn-Tucker é que as irregularidades que mencionamos anteriormente não dizem respeito à forma da região viável por si, mas às formas das próprias funções de restrição.

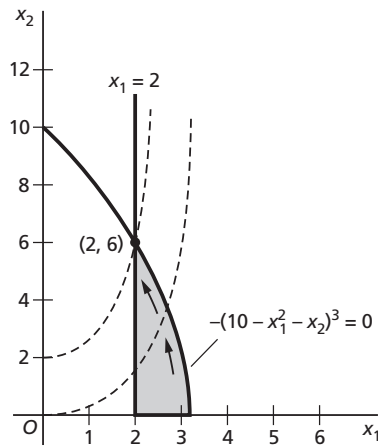


FIGURA 13.3

A qualificação de restrição

Irregularidades de fronteira – com ou sem cúspides – não ocorrerão se for satisfeita uma certa qualificação de restrição.

Para explicar isso, seja $x^* \equiv (x_1^*, x_2^*, \dots, x_n^*)$ um ponto de fronteira da região viável e um possível candidato a uma solução, e suponha que $dx \equiv (dx_1, dx_2, \dots, dx_n)$ representa a direção de um movimento específico do citado ponto de fronteira. A interpretação do vetor dx como representante da direção do movimento está perfeitamente de acordo com nossa interpretação anterior de um vetor como um segmento de reta direcionado (uma seta), porém, aqui, o ponto de partida é o ponto x^* em vez do ponto de origem e, portanto, o vetor dx não tem as características de um vetor raio. Agora vamos impor dois requisitos ao vetor dx . Primeiro, se a j -ésima variável de escolha tiver um valor zero no ponto x^* , então permitiremos somente uma mudança não-negativa no eixo x_j ; isto é,

$$dx_j \geq 0 \quad \text{se} \quad x_j^* = 0 \quad (13.22)$$

Segundo, se a i -ésima restrição for exatamente satisfeita no ponto x^* , então permitiremos somente valores de $dx_1, \dots, dx_n$ tais que o valor da função de restrição $g^i(x^*)$ não aumente (para um problema de maximização) nem diminua (para um problema de minimização), isto é,

$$dg^i(x^*) = g_1^i dx_1 + g_2^i dx_2 + \dots + g_n^i dx_n \begin{cases} \leq 0 (\text{máximo}) \\ \geq 0 (\text{mínimo}) \end{cases} \quad \text{se} \quad g^i(x^*) = r_i \quad (13.23)$$

onde todas as derivadas parciais de g_j^i devem ser calculadas em x^* . Se um vetor dx satisfizer (13.22) e (13.23), nos referiremos a ele como um *vetor teste*. Por fim, se existir um arco diferenciável que (1) parte do ponto x^* ; (2) é contido inteiramente na região viável; e (3) é tangente a um dado vetor teste, nos o denominaremos um *arco qualificador* para aquele vetor teste. Tendo isso como base, a qualificação da restrição pode ser enunciada simplesmente como se segue:

A qualificação de restrição é satisfeita se para qualquer ponto x^* na fronteira da região viável existir um arco qualificador para cada vetor teste dx .

EXEMPLO 4

Mostraremos que o ponto ótimo $(1, 0)$ do Exemplo 1 na Figura 13.2, que não atende às condições de Kuhn-Tucker, também não cumpre a qualificação de restrição. Naquele ponto, $x_2^* = 0$; assim, o vetor teste deve satisfazer

$$dx_2 \geq 0 \quad [\text{por (13.22)}]$$

Além disso, visto que a (única) restrição, $g^1 = x_2 - (1 - x_1)^3 \leq 0$, é exatamente satisfeita em $(1, 0)$, temos de ter [por (13.23)]

$$g_1^1 dx_1 + g_2^1 dx_2 = 3(1 - x_1)^2 dx_1 + dx_2 = dx_2 \leq 0$$

Juntos, esses dois requisitos implicam que temos de ter $dx_2 = 0$. Por outro lado, somos livres para escolher dx_1 . Assim, por exemplo, o vetor $(dx_1, dx_2) = (2, 0)$ é um vetor teste aceitável, bem como $(dx_1, dx_2) = (-1, 0)$. Este último vetor teste seria representado graficamente na Figura 13.2 como uma seta que parte de $(1, 0)$ e aponta para oeste (não desenhada) e é claramente possível desenhar um arco qualificador para ele. (A própria fronteira curvada da região viável pode servir como um arco qualificador.) Por outro lado, o vetor teste $(dx_1, dx_2) = (2, 0)$ seria representado graficamente por uma seta partindo de $(1, 0)$ e apontando para leste (não desenhada). Uma vez que não há nenhum modo de desenhar um arco suave tangente a esse vetor e que fique inteiramente dentro da região viável, não existe nenhum arco qualificador para ele. Por conseguinte, o ponto de solução ótima $(1, 0)$ viola a qualificação de restrição.

EXEMPLO 5

Com referência ao Exemplo 2, vamos ilustrar que, após ser acrescentada uma restrição adicional $2x_1 + x_2 \leq 2$ à Figura 13.2, o ponto $(1, 0)$ satisfará a qualificação de restrição, revalidando, desse modo, as condições de Kuhn-Tucker.

Como no Exemplo 4, temos de exigir $dx_2 \geq 0$ (porque $x_2^* = 0$) e $dx_2 \leq 0$ (porque a primeira restrição é exatamente satisfeita); assim, $dx_2 = 0$. Mas a segunda restrição também é exatamente satisfeita e, por isso, requer

$$g_1^2 dx_1 + g_2^2 dx_2 = 2dx_1 + dx_2 = 2dx_1 \leq 0 \quad [\text{por (13.23)}]$$

Com dx_1 não-positiva e dx_2 zero, os únicos vetores teste admissíveis – fora o próprio vetor nulo – são os vetores que apontam para oeste na Figura 13.2 e partem de $(1, 0)$. Todos esses vetores estão ao longo do eixo horizontal na região viável e certamente é possível desenhar um arco qualificador para cada vetor teste. Por conseguinte, desta vez, a qualificação de restrição é, de fato, satisfeita.

Restrições lineares

No Exemplo 3 anterior, foi demonstrado que a convexidade do conjunto viável não garante a validade das condições de Kuhn-Tucker como condições necessárias. Contudo, se a região viável for um conjunto convexo formado somente por restrições *lineares*, então a qualificação de restrição será invariavelmente cumprida e as condições de Kuhn-Tucker sempre serão válidas em uma solução ótima. Sendo esse o caso, nunca precisaremos nos incomodar com as irregularidades da fronteira quando se tratar de um problema de programação não-linear com restrições lineares.

EXEMPLO 6 Vamos ilustrar o resultado da restrição linear na estrutura de duas variáveis, duas restrições. Para um problema de maximização, as restrições lineares podem ser escritas como

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &\leq r_1 \\ a_{21}x_1 + a_{22}x_2 &\leq r_2 \end{aligned}$$

onde tomaremos todos os parâmetros como positivos. Então, com o indicado na Figura 13.4, a primeira fronteira da restrição terá uma inclinação de $-a_{11}/a_{12} < 0$, e a segunda, uma inclinação de $-a_{21}/a_{22} < 0$. Os pontos de fronteira da região viável podem ser de um dos cinco tipos seguintes: (1) o ponto de origem, onde os dois eixos se interceptam; (2) pontos que estão sobre um segmento de eixo, tais como J e S ; (3) pontos na interseção de um eixo e de uma fronteira de restrição, a saber, K e R ; (4) pontos que estão sobre uma única fronteira de restrição, como L e N ; e, (5) o ponto de interseção de duas restrições, M . Vamos examinar brevemente cada tipo por vez, no que diz respeito à satisfação da qualificação de restrição.

1. Na origem, nenhuma restrição é exatamente satisfeita, portanto, podemos ignorar (13.23). Mas, visto que $x_1 = x_2 = 0$, devemos escolher vetores teste com $dx_1 \geq 0$ e $dx_2 \geq 0$, por (13.22). Por conseguinte, todos os vetores teste que partem da origem apontam para leste, norte ou nordeste, como ilustrado na Figura 13.4. A propósito, todos esses vetores caem dentro do conjunto viável, e fica bem claro que é possível encontrar um arco qualificador para cada um deles.
2. Em um ponto como J , novamente podemos ignorar (13.23). O fato de que $x_2 = 0$ significa que devemos escolher $dx_2 \geq 0$, mas podemos escolher dx_1 livremente. Por conseguinte, todos os vetores seriam aceitáveis, exceto os que apontam para o sul ($dx_2 < 0$). Mais uma vez, todos esses vetores caem dentro da região viável e existe um arco qualificador para cada um deles. A análise do ponto S é semelhante.
3. Nos pontos K e R , as duas, (13.22) e (13.23), devem ser consideradas. Especificamente, em K , temos de escolher $dx_2 \geq 0$, já que $x_2 = 0$, de modo que devemos excluir todas as setas que apontam para o sul. Além do mais, como a segunda restrição é exatamente satisfeita, os vetores teste para o ponto K têm de satisfazer

$$g_1^2 dx_1 + g_2^2 dx_2 = a_{21} dx_1 + a_{22} dx_2 \leq 0 \quad (13.24)$$

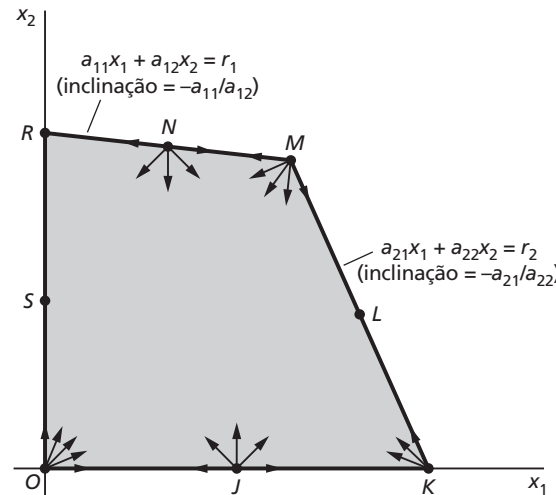


FIGURA 13.4

Contudo, visto que em K também temos $a_{21}x_1 + a_{22}x_2 = r_2$ (segunda fronteira de restrição), podemos acrescentar essa igualdade a (13.24) e modificar a restrição ao vetor teste para a forma

$$a_{21}(x_1 + dx_1) + a_{22}(x_2 + dx_2) \leq r_2 \quad (13.24')$$

Interpretando $(x_j + dx_j)$ como o novo valor de x_j atingido na ponta da seta de um vetor teste, podemos inferir que (13.24') significa que as pontas das setas de todos os vetores teste devem estar localizadas sobre ou abaixo da segunda fronteira de restrição. Por consequência, todos esses vetores novamente devem cair dentro da região viável, e é possível encontrar um arco qualificador para cada um. A análise do ponto R é análoga.

4. Em pontos como L e N , nenhuma das variáveis é zero e (13.22) pode ser ignorada. Entretanto, para o ponto N , (13.23) determina que

$$g_1^1 dx_1 + g_2^1 dx_2 = a_{11} dx_1 + a_{12} dx_2 \leq 0 \quad (13.25)$$

Visto que o ponto N satisfaz $a_{11} dx_1 + a_{12} dx_2 = r_1$ (primeira fronteira de restrição), podemos acrescentar essa igualdade a (13.25) e escrever

$$a_{11}(x_1 + dx_1) + a_{12}(x_2 + dx_2) \leq r_1 \quad (13.25')$$

Isso exigiria que as pontas das setas dos vetores teste estivessem localizadas sobre ou abaixo da primeira fronteira de restrição na Figura 13.4. Assim, obtemos essencialmente o mesmo tipo de resultado encontrado nos outros casos. Essa análise é análoga no ponto L .

5. No ponto M podemos novamente desprezar (13.22), mas, desta vez, (13.23) requer que todos os vetores teste satisfaçam tanto (13.24) quanto (13.25). Uma vez que podemos modificar estas últimas condições para as formas em (13.24') e (13.25'), as pontas das setas de todos os vetores teste agora devem estar localizadas sobre ou abaixo da primeira, bem como da segunda fronteira de restrição. Assim, mais uma vez o resultado reproduz os dos casos anteriores.

A propósito, neste exemplo, para cada tipo de ponto de fronteira considerado, todos os vetores teste caem dentro da região viável. Conquanto essa característica de localização facilite achar os arcos qualificadores, ela não é, de modo algum, um pré-requisito da existência desses arcos. Quando o problema se referir a uma fronteira de restrição não-linear, em particular, a própria fronteira de restrição pode servir como um arco qualificador para algum vetor teste localizado fora da região viável. Um exemplo disso pode ser encontrado em um dos problemas a seguir.

EXERCÍCIO 13.2

- Verifique se o ponto de solução $(x_1^*, x_2^*) = (2, 6)$ no Exemplo 3 satisfaz a qualificação de restrição.
- Maximize $\pi = x_1$
sujeita a $x_1^2 + x_2^2 \leq 1$
e $x_1, x_2 \geq 0$
Resolva graficamente e verifique se o ponto de solução ótima satisfaz (a) a qualificação de restrição e (b) as condições de Kuhn-Tucker.
- Minimize $C = x_1$
sujeita a $x_1^2 - x_2 \geq 0$
e $x_1, x_2 \geq 0$
Resolva graficamente. A solução ótima ocorre em uma cúspide? Verifique se a solução ótima satisfaz (a) a qualificação de restrição e (b) as condições de Kuhn-Tucker para mínimo.

4. Minimize $C = x_1$
sujeita a $-x_2 - (1 - x_1)^3 \geq 0$
e $x_1, x_2 \geq 0$

Mostre que (a) a solução ótima $(x_1^*, x_2^*) = (1, 0)$ não satisfaz as condições de Kuhn-Tucker, mas (b) que, introduzindo um novo multiplicador $\lambda_0 \geq 0$, e modificando a função de Lagrange (13.15) para a forma

$$Z_0 = \lambda_0 f(x_1, x_2, \dots, x_n) + \sum_{i=1}^m \lambda_i [r_i - g^i(x_1, x_2, \dots, x_n)]$$

as condições de Kuhn-Tucker podem ser satisfeitas em $(1, 0)$. (Nota: As condições de Kuhn-Tucker sobre os multiplicadores se estendem somente a $\lambda_1, \dots, \lambda_m$, mas não a λ_0 .)

13.3 Aplicações econômicas

Racionamento em tempo de guerra

Durante tempos de guerra, é típico que a população civil seja sujeita a algum tipo de racionamento de bens de consumo básicos. Em geral, o método de racionamento é a utilização de cupons resgatáveis emitidos pelo governo. O governo fornecerá a cada consumidor uma quota de cupons por mês. O consumidor, por sua vez, terá de resgatar um certo número de cupons na ocasião da compra de um bem racionado. Isso significa que, em essência, o consumidor paga *dois* preços na hora da compra. Ele paga o preço do cupom, bem como o preço monetário do bem racionado, o que exige que o consumidor tenha fundos suficientes e também cupons suficientes para comprar uma unidade do bem racionado.

Considere o caso de um mundo onde existam apenas dois bens, x e y , ambos racionados. Seja $U = U(x, y)$ a função utilidade do consumidor. Ele tem um orçamento monetário fixo B e enfrenta preços exógenos P_x e P_y . Além disso, o consumidor tem uma quota de cupons denotada por C , que pode ser usada para comprar ou x ou y a um preço de cupom de c_x e c_y . Portanto, o problema de maximização do consumidor é

$$\begin{array}{ll} \text{Maximize} & U = U(x, y) \\ \text{sujeita a} & P_x x + P_y y \leq B \\ & c_x x + c_y y \leq C \\ \text{e} & x, y \geq 0 \end{array}$$

A função de Lagrange para o problema é

$$Z = U(x, y) + \lambda_1 (B - P_x x - P_y y) + \lambda_2 (C - c_x x - c_y y)$$

onde λ_1 e λ_2 são os multiplicadores de Lagrange. Posto que ambas as restrições são lineares, a qualificação de restrição é satisfeita e as condições de Kuhn-Tucker são necessárias:

$$\begin{array}{lll} Z_x = U_x - \lambda_1 P_x - \lambda_2 c_x \leq 0 & x \geq 0 & x Z_x = 0 \\ Z_y = U_y - \lambda_1 P_y - \lambda_2 c_y \leq 0 & y \geq 0 & y Z_y = 0 \\ Z_{\lambda_1} = B - P_x x - P_y y \geq 0 & \lambda_1 \geq 0 & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = C - c_x x - c_y y \geq 0 & \lambda_2 \geq 0 & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

EXEMPLO 1 Suponha que a função utilidade é da forma $U = xy^2$. Além disso, seja $B = 100$ e $P_x = P_y = 1$ enquanto $C = 120$, $c_x = 2$ e $c_y = 1$.

A função de Lagrange toma a forma específica de

$$Z = xy^2 + \lambda_1(100 - x - y) + \lambda_2(120 - 2x - y)$$

As condições de Kuhn-Tucker agora são

$$\begin{array}{lll} Z_x = y^2 - \lambda_1 - 2\lambda_2 \leq 0 & x \geq 0 & xZ_x = 0 \\ Z_y = 2xy - \lambda_1 - \lambda_2 \leq 0 & y \geq 0 & yZ_y = 0 \\ Z_{\lambda_1} = 100 - x - y \geq 0 & \lambda_1 \geq 0 & \lambda_1 Z_{\lambda_1} = 0 \\ Z_{\lambda_2} = 120 - 2x - y \geq 0 & \lambda_2 \geq 0 & \lambda_2 Z_{\lambda_2} = 0 \end{array}$$

Mais uma vez, o procedimento de solução envolve uma certa quantidade de tentativa e erro. Primeiro, podemos escolher que uma das restrições seja não-vinculadora e resolver para x e y . Uma vez encontrados esses valores, usá-los para testar se a restrição escolhida como não-vinculadora é violada. Se for, então repetimos o procedimento escolhendo uma outra restrição como não-vinculadora. Se ocorrer novamente a violação da restrição não-vinculadora, podemos admitir que ambas as restrições são vinculadoras e a solução é determinada somente pelas restrições.

Etapa 1: Suponha que a segunda restrição (acionamento) seja não-vinculadora na solução, de modo que $\lambda_2 = 0$ por folga complementar. Mas sejam x , y e λ_1 positivos, de modo que a folga complementar nos daria as três equações seguintes:

$$\begin{array}{l} Z_x = y^2 - \lambda_1 = 0 \\ Z_y = 2xy - \lambda_1 = 0 \\ Z_{\lambda_1} = 100 - x - y = 0 \end{array}$$

Resolvendo para x e y , temos uma solução experimental

$$x = 33\frac{1}{3} \quad y = 66\frac{2}{3}$$

Contudo, quando substituímos essas soluções na restrição de cupom, constatamos que

$$2(33\frac{1}{3}) + 66\frac{2}{3} = 133\frac{1}{3} > 120$$

Essa solução viola a restrição de cupom e deve ser rejeitada.

Etapa 2: Agora vamos inverter as premissas em λ_1 e λ_2 de modo que $\lambda_1 = 0$, mas λ_2 , x , $y > 0$. Então, pelas condições marginais, temos

$$\begin{array}{l} Z_x = y^2 - 2\lambda_2 = 0 \\ Z_y = 2xy - \lambda_2 = 0 \\ Z_{\lambda_2} = 120 - 2x - y = 0 \end{array}$$

Resolver esse sistema de equações resulta em uma outra solução experimental

$$x = 20 \quad y = 80$$

que implica que $\lambda_2 = 2xy = 3.200$. Esses valores de solução, juntamente com $\lambda_1 = 0$, satisfazem ambas as restrições, a orçamentária e a de racionamento. Assim, podemos aceitá-las como a solução final para as condições de Kuhn-Tucker.

Contudo, essa solução ótima contém uma anormalidade curiosa. Como a restrição orçamentária é vinculadora na solução, normalmente esperaríamos que o multiplicador de Lagrange relacionado fosse positivo, mas, na verdade, temos $\lambda_1 = 0$. Assim, neste exemplo, conquanto a restrição orçamentária seja *matematicamente* vinculadora (satisfeita como uma igualdade restrita na solução), ela é *economicamente* não-vinculadora (não exige uma utilidade monetária marginal positiva).

Determinação de preço de pico (carga máxima)

A determinação de preço de pico (carga máxima) e de preço fora do pico e problemas de planejamento são questões corriqueiras para empresas cujos processos de produção estão sujeitos a restrições de capacidade. Usualmente, a empresa investiu em capacidade de modo a atingir um mercado primário. Todavia, pode existir um mercado secundário onde a empresa possa vender seu produto. Uma vez comprado o equipamento principal para atender o mercado primário da empresa, ele estará livremente disponível (até o limite de sua capacidade) para ser usado no mercado secundário. Alguns exemplos típicos são escolas e universidades que constroem suas instalações para suprir as necessidades do período diurno (pico), mas podem oferecer cursos noturnos (fora do pico); teatros que oferecem espetáculos à noite (pico) e à tarde (fora do pico); e empresas transportadoras rodoviárias cujas rotas são dedicadas, mas que podem optar por entrar em mercados de “carga de retorno”. Visto que o custo de capacidade é um fator na decisão de maximização de lucro para o mercado de pico e já está pago, ele normalmente não seria um fator no cálculo do preço e quantidades ótimas para o mercado menor, fora do pico. Contudo, se a demanda do mercado secundário for próxima ou do mesmo tamanho da demanda do mercado primário, restrições de capacidade podem se tornar uma questão importante, especialmente porque discriminar e cobrar preços mais baixos em períodos fora do pico é uma prática comum. Mesmo que o mercado secundário seja menor que o primário, é possível que, ao preço mais baixo (maximizado de lucro), a demanda fora do pico exceda a capacidade. Nesses casos, a escolha das capacidades deve levar ambos os mercados em conta, o que transforma o problema em uma aplicação clássica de programação não-linear.

Considere uma empresa maximizadora de lucro que enfrenta duas curvas de receita média

$$P_1 = P^1(Q_1) \text{ no período diurno (período de pico)}$$

$$P_2 = P^2(Q_2) \text{ no período noturno (período fora do pico)}$$

Para funcionar, a empresa tem de pagar b por unidade de produto, seja de dia, seja à noite. Além disso, tem de comprar capacidade a um custo de c por unidade de capacidade. Vamos denotar a capacidade total por K , medida em unidades de Q . A empresa deve pagar por capacidade mesmo que funcione no período fora de pico. De quem seriam cobrados os custos de capacidade: do conjunto de clientes de pico, fora do pico ou ambos? O problema de maximização da empresa se torna

$$\text{Maximize}_{Q_1, Q_2, K} \quad \pi = P_1 Q_1 + P_2 Q_2 - b(Q_1 + Q_2) - cK$$

$$\text{sujeita a} \quad \begin{aligned} Q_1 &\leq K \\ Q_2 &\leq K \end{aligned}$$

$$\text{onde} \quad \begin{aligned} P_1 &= P^1(Q_1) \\ P_2 &= P^2(Q_2) \end{aligned}$$

$$\text{e} \quad Q_1, Q_2, K \geq 0$$

Em vista de a receita total para Q_i

$$R_i \equiv P_i Q_i = P^i(Q_i) Q_i$$

ser uma função de Q_i apenas, podemos simplificar o enunciado do problema para

$$\begin{array}{ll}
 \text{Maximize} & \pi = R_1(Q_1) + R_2(Q_2) - b(Q_1 + Q_2) - cK \\
 \text{sujeita a} & Q_1 \leq K \\
 & Q_2 \leq K \\
 \text{e} & Q_1, Q_2, K \geq 0
 \end{array}$$

Note que ambas as restrições são lineares; assim, a qualificação de restrição é satisfeita e as condições de Kuhn-Tucker são necessárias.

A função de Lagrange é

$$Z = R_1(Q_1) + R_2(Q_2) - b(Q_1 + Q_2) - cK + \lambda_1(K - Q_1) + \lambda_2(K - Q_2)$$

e as condições de Kuhn-Tucker são

$$\begin{array}{lll}
 Z_1 = MR_1 - b - \lambda_1 \leq 0 & Q_1 \geq 0 & Q_1 Z_1 = 0 \\
 Z_2 = MR_2 - b - \lambda_2 \leq 0 & Q_2 \geq 0 & Q_2 Z_2 = 0 \\
 Z_K = -c + \lambda_1 + \lambda_2 \leq 0 & K \geq 0 & K Z_K = 0 \\
 Z_{\lambda_1} = K - Q_1 \geq 0 & \lambda_1 \geq 0 & \lambda_1 Z_{\lambda_1} = 0 \\
 Z_{\lambda_2} = K - Q_2 \geq 0 & \lambda_2 \geq 0 & \lambda_2 Z_{\lambda_2} = 0
 \end{array}$$

onde MR_i é a receita marginal de Q_i ($i = 1, 2$).

O procedimento de solução novamente acarreta tentativa e erro. Vamos admitir, em primeiro lugar, que $Q_1, Q_2, K > 0$. Então, por folga complementar, temos

$$\begin{aligned}
 MR_1 - b - \lambda_1 &= 0 \\
 MR_2 - b - \lambda_2 &= 0 \\
 -c + \lambda_1 + \lambda_2 &= 0 \quad (\lambda_1 = c - \lambda_2)
 \end{aligned} \tag{13.26}$$

que podem ser condensadas em duas equações após eliminar λ_1 :

$$\begin{aligned}
 MR_1 &= b + c - \lambda_2 \\
 MR_2 &= b + \lambda_2
 \end{aligned} \tag{13.26'}$$

Então, prosseguimos em duas etapas.

Etapas 1: Visto que o mercado fora do pico é um mercado secundário, podemos esperar que sua função de receita marginal (MR_2) esteja abaixo da receita marginal do mercado primário (MR_1), como ilustrado na Figura 13.5. Além disso, é mais provável que a restrição de capacidade seja não-vinculadora no mercado secundário, de modo que λ_2 tem maior probabilidade de ser. Portanto, experimentamos $\lambda_2 = 0$. Então (13.26') se torna

$$\begin{aligned}
 MR_1 &= b + c \\
 MR_2 &= b
 \end{aligned} \tag{13.26''}$$

O fato de o mercado primário absorver todo o custo de capacidade c implica que $Q_1 = K$. Contudo, ainda precisamos verificar se a restrição $Q_2 \leq K$ é satisfeita. Se for, a solução que encontramos é válida. A Figura 13.5(a) ilustra o caso em que $Q_1 = K$ e $Q_2 < K$ na solução. A curva MR_1 intercepta a reta $b + c$ no ponto E_1 e a curva MR_2 intercepta a reta b no ponto E_2 .

E se a solução experimental anterior acarretar $Q_2 > K$, como ocorreria se a curva MR_2 estivesse muito próxima de MR_1 , de modo a interceptar a reta b em uma produção maior do que K ?

Então, é claro que a segunda restrição é violada, e devemos rejeitar a premissa $\lambda_2 = 0$ e passar para a próxima etapa.

Etapa 2: Agora, vamos admitir que ambos os multiplicadores de Lagrange são positivos e, assim, $Q_1 = Q_2 = K$. Então, como não podemos eliminar qualquer variável de (13.26), temos

$$\begin{aligned} MR_1 &= b + \lambda_1 \\ MR_2 &= b + \lambda_2 \\ c &= \lambda_1 + \lambda_2 \end{aligned} \quad (13.26''')$$

Esse caso está ilustrado na Figura 13.5(b), onde os pontos E_1 e E_2 satisfazem as duas primeiras equações em (13.26'''). Pela terceira equação, vemos que o custo de capacidade c é a soma dos dois multiplicadores de Lagrange. Isso significa que λ_1 e λ_2 representam as porções do custo de capacidade suportadas respectivamente pelos dois mercados.

EXEMPLO 2 Suponha que a função de receita média durante as horas de pico é

$$P_1 = 22 - 10^{-5}Q_1$$

e que, durante as horas fora do pico é

$$P_2 = 18 - 10^{-5}Q_2$$

Produzir uma unidade de produto a cada meio dia requer uma unidade de capacidade que custa 8 centavos por dia. O custo de uma unidade de capacidade é o mesmo quer a capacidade seja usada apenas nos horários de pico, quer nos horários fora de pico também. Produzir uma unidade por meio-dia custa, além dos custos de capacidade, 6 centavos de custos operacionais (trabalho mais combustível) – de dia e de noite.

Se admitirmos que a restrição de capacidade é não-vinculadora no mercado secundário ($\lambda_2 = 0$), então as condições de Kuhn-Tucker dadas tornam-se

$$\begin{aligned} \lambda_1 &= c = 8 \\ 22 - 2 \times 10^{-5}Q_1 &= b + c = 14 \\ 18 - 2 \times 10^{-5}Q_2 &= b = 6 \end{aligned}$$

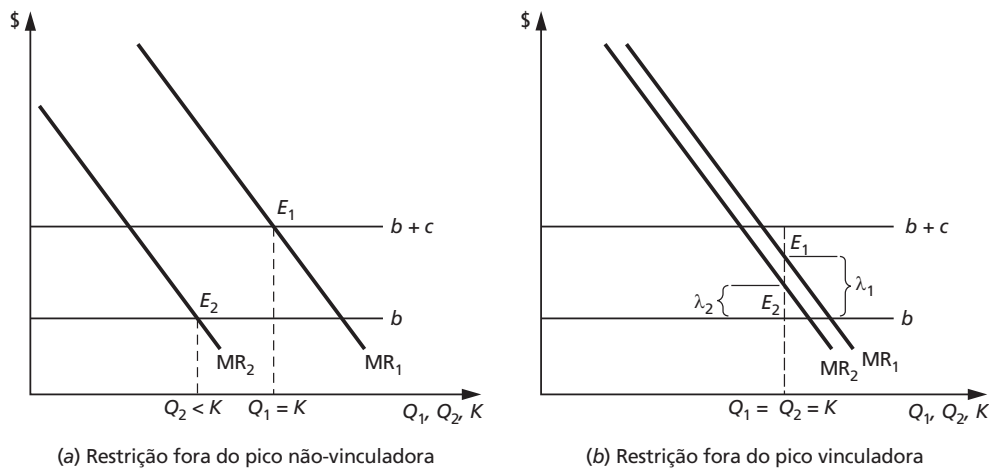


FIGURA 13.5

A resolução desse sistema nos dá

$$Q_1 = 400.000$$

$$Q_2 = 600.000$$

o que viola a premissa adotada, isto é, a segunda restrição é não-vinculadora, porque $Q_2 > Q_1 = K$.

Portanto, vamos admitir que ambas as restrições são vinculadoras. Então $Q_1 = Q_2 = Q$ e as condições de Kuhn-Tucker tornam-se

$$\lambda_1 + \lambda_2 = 8$$

$$22 - 2 \times 10^{-5}Q = 6 + \lambda_1$$

$$18 - 2 \times 10^{-5}Q = 6 + \lambda_2$$

que dá a seguinte solução

$$Q_1 = Q_2 = K = 500.000$$

$$\lambda_1 = 6 \quad \lambda_2 = 2$$

$$P_1 = 17 \quad P_2 = 13$$

Uma vez que a restrição de capacidade é vinculadora em ambos os mercados, o mercado primário paga $\lambda_1 = 6$ do custo de capacidade e o mercado secundário paga $\lambda_2 = 2$.

EXERCÍCIO 13.3

- Suponha que, no Exemplo 2, uma unidade de capacidade custa somente 3 centavos por dia.
 - Quais seriam os preços e quantidades de pico e fora do pico que maximizariam o lucro?
 - Quais seriam os valores dos multiplicadores de Lagrange? Como você interpretaria esses valores?
- Uma consumidora mora em uma ilha onde ela produz dois bens, x e y , de acordo com o limite de possibilidade de produção $x^2 + y^2 \leq 200$, e ela mesma consome todos os bens. Sua função utilidade é $U = xy^3$.
A consumidora também enfrenta uma restrição ambiental imposta à produção total de ambos os bens. A restrição ambiental é dada por $x + y \leq 20$.
 - Escreva as condições de primeira ordem de Kuhn-Tucker.
 - Ache os x e y ótimos da consumidora. Identifique quais restrições são vinculadoras.
- Uma empresa fornecedora de energia elétrica está construindo uma estação em um país estrangeiro e tem de planejar sua capacidade. A demanda de energia no período de pico é dada por $P_1 = 400 - Q_1$ e a demanda fora do período de pico é dada por $P_2 = 380 - Q_2$. O custo variável é 20 por unidade (pago em ambos os mercados) e a capacidade custa 10 por unidade e é paga somente uma vez e usada em ambos os períodos.
 - Escreva a função de Lagrange e as condições de Kuhn-Tucker para esse problema.
 - Ache as produções e capacidade ótimas para esse problema.
 - Que porção da capacidade é paga por mercado (isto é, quais são os valores de λ_1 e λ_2)?
 - Agora suponha que o custo de capacidade é 30 centavos por unidade (pago somente uma vez). Ache quantidades, capacidade e a porção da capacidade paga por mercado (isto é, λ_1 e λ_2).

13.4 Teoremas de suficiência em programação não-linear

Nas seções anteriores, introduzimos as condições de Kuhn-Tucker e ilustramos suas aplicações como condições *necessárias* em problemas de otimização com restrições de desigualdade. Sob certas circunstâncias, as condições de Kuhn-Tucker também podem ser tomadas como condições suficientes.

O teorema de suficiência de Kuhn-Tucker: programação côncava

Em problemas clássicos de otimização, as condições suficientes para máximo e mínimo são tradicionalmente expressas em termos dos sinais de derivadas ou diferenciais de segunda ordem. Como mostramos na Seção 11.5, entretanto, essas condições de segunda ordem estão estreitamente relacionadas com os conceitos de concavidade e convexidade da função objetivo. Aqui, em programação não-linear, as condições suficientes também podem ser enunciadas diretamente em termos de concavidade e convexidade. E, de fato, esses conceitos serão aplicados não somente à função objetivo $f(x)$, mas também às funções restrição $g^i(x)$.

Para o problema da *maximização*, Kuhn e Tucker oferecem o seguinte enunciado para condições suficientes (teorema de suficiência):

Dado o problema de programação não-linear

$$\begin{array}{ll} \text{Maximize} & \pi = f(x) \\ \text{sujeita a} & g^i(x) \leq r_i \quad (i = 1, 2, \dots, m) \\ \text{e} & x \geq 0 \end{array}$$

se as seguintes condições forem satisfeitas:

- (a) a função objetivo $f(x)$ é diferenciável e *côncava* no ortante não-negativo
- (b) cada função de restrição $g^i(x)$ é diferenciável e *convexa* no ortante não-negativo
- (c) o ponto x^* satisfaz as condições de máximo de Kuhn-Tucker

então x^* dá um máximo global de $\pi = f(x)$.

Note que, nesse teorema, a qualificação de restrição não é mencionada em lugar nenhum. Isso porque na condição (c) já supusemos que as condições de Kuhn-Tucker são satisfeitas em x^* e, por conseqüência, a questão da qualificação de restrição deixa de ter importância.

Como enunciado, o teorema apresentado indica que as condições (a), (b) e (c) são suficientes para estabelecer x^* como uma solução ótima. Porém, olhando com outros olhos, uma outra interpretação possível é que ele significa que, dados (a) e (b), então as condições de Kuhn-Tucker para máximo são suficientes para um máximo. Na seção anterior, aprendemos que as condições de Kuhn-Tucker, embora não necessárias *por si*, tornam-se necessárias quando a qualificação de restrição é satisfeita. Combinando essa informação com o teorema de suficiência, agora podemos afirmar que, se a qualificação de restrição for satisfeita e se as condições (a) e (b) forem realizadas, então as condições de Kuhn-Tucker para máximo serão *necessárias e suficientes* para um máximo. Este seria o caso, por exemplo, quando todas as restrições são desigualdades lineares, as quais são suficientes para satisfazer a qualificação de restrição.

O problema de maximização de que tratamos no teorema de suficiência citado é freqüentemente denominado *programação côncava*. Esse nome aparece porque Kuhn e Tucker adotam a desigualdade $\geq$ em vez da desigualdade $\leq$ em cada restrição, de modo que a condição (b) exigiria que as funções $g^i(x)$ fossem *todas côncavas*, como a função $f(x)$. Porém, modificamos a formulação para transmitir a idéia de que, em um problema de maximização, uma restrição é imposta para “frear” (daí, $\leq$) a tentativa de subir a pontos mais altos na função objetivo. Embora as duas formulações sejam diferentes na forma, elas são equivalentes em substância. Por questão de brevidade, omitimos a demonstração.

Como já mencionado, o teorema de suficiência trata somente de problemas de maximização. Mas a adaptação para problemas de *minimização* não é, de modo algum, difícil. À parte as modificações apropriadas no teorema de modo a refletir a inversão do problema em si, basta per-

mutar as duas palavras *côncava* e *convexa* nas condições (a) e (b) e usar as condições de Kuhn-Tucker para *mínimo* na condição (c). (Veja Exercício 13.4–1.)

O teorema de suficiência de Arrow-Enthoven: programação quase-côncava

Para aplicar o teorema de suficiência de Kuhn-Tucker, é preciso cumprir certas especificações de concavidade-convexidade. Essas especificações constituem requisitos bastante rigorosos. Em um outro teorema de suficiência – o teorema de suficiência de Arrow-Enthoven[†] – essas especificações são afrouxadas até o ponto de exigir somente *quase-concavidade* e *quase-convexidade* nas funções objetivo e restrição. Com esse abrandamento dos requisitos, o escopo de aplicabilidade das condições suficientes também é proporcionalmente ampliado.

Na formulação original apresentada no artigo de Arrow-Enthoven, quando temos um problema de maximização com restrições na forma $\geq$, as funções $f(x)$ e $g^i(x)$ devem ser uniformemente quase-côncavas para que seu teorema seja aplicável. Isso dá origem ao nome *programação quase-côncava*. Contudo, na discussão que faremos aqui, usaremos novamente a desigualdade $\leq$ nas restrições de um problema de maximização e a desigualdade $\geq$ no problema de minimização.

O teorema é o seguinte:

Dado o problema de programação não-linear

$$\begin{array}{ll} \text{Maximize} & \pi = f(x) \\ \text{sujeita a} & g^i(x) \leq r_i \quad (i = 1, 2, \dots, m) \\ \text{e} & x \geq 0 \end{array}$$

se as seguintes condições forem satisfeitas:

- (a) a função objetivo $f(x)$ é diferenciável e *quase-côncava* no ortante não-negativo
- (b) cada função restrição $g^i(x)$ é diferenciável e *quase-convexa* no ortante não-negativo
- (c) o ponto x^* satisfaz as condições de Kuhn-Tucker para máximo
- (d) qualquer *uma* das seguintes é satisfeita:
 - (d-i) $f_j(x^*) < 0$ para, no mínimo, uma variável x_j
 - (d-ii) $f_j(x^*) > 0$ para alguma variável x_j que possa assumir um valor positivo sem violar as restrições
 - (d-iii) as n derivadas $f_j(x^*)$ não são todas zero e a função $f(x)$ é duas vezes diferenciável na vizinhança de x^* [isto é, todas as derivadas parciais de segunda ordem de $f(x)$ existem em x^*]
 - (d-iv) a função $f(x)$ é côncava

então x^* dá um máximo global de $\pi = f(x)$.

Uma vez que a demonstração desse teorema é bastante longa, vamos omiti-la aqui. Contudo, queremos chamar sua atenção para alguns aspectos importantes desse teorema. Para começar, conquanto Arrow e Enthoven tenham conseguido enfraquecer as especificações de concavidade-convexidade para suas contrapartes quase-concavidade-quase-convexidade, eles acham necessário acrescentar um novo requisito, (d). Porém, note que exige-se que somente *uma* das quatro alternativas listadas em (d) forme um conjunto completo de condições suficientes. Por conseguinte, na verdade o teorema acima contém um número total de *quatro* conjuntos diferentes de condições suficientes para um máximo. No caso de (d-iv), com $f(x)$ côncava, aparentemente o teorema de suficiência de Arrow-Enthoven se tornaria idêntico ao teorema de suficiência de Kuhn-Tucker. Mas isso não é verdade. Considerando que Arrow e Enthoven exigem somente que as funções de restrição $g^i(x)$ sejam *quase-convexas*, suas condições suficientes são ainda mais fracas.

Como enunciado, o teorema engloba as condições de (a) a (d) em um conjunto de condições suficientes. Mas também é possível interpretar que ele significa que, quando (a), (b) e (d)

[†] Arrow, Kenneth J. e Enthoven, Alain C. "Quasi-concave Programming". *Econometrica*, p. 779–800, out. 1961.

forem satisfeitas, as condições de Kuhn-Tucker para máximo se tornarão condições suficientes para um máximo. Além do mais, se a qualificação de restrição também for satisfeita, as condições de Kuhn-Tucker se tornarão necessárias e suficientes para um máximo.

Como acontece com o teorema de Kuhn-Tucker, o teorema de Arrow-Enthoven pode ser adaptado com facilidade para a estrutura de *minimização*. À parte as modificações óbvias que são necessárias para reverter a direção de otimização, basta permutar as palavras *quase-côncava* e *quase-convexa* nas condições (a) e (b), substituir as condições de Kuhn-Tucker para máximo pelas condições para mínimo, inverter as desigualdades em (d-i) e (d-ii), e trocar a palavra *côncava* por *convexa* em (d-iv).

Um teste de qualificação de restrição

Foi mencionado na Seção 13.2 que, se todas as funções de restrição forem lineares, a qualificação de restrição é satisfeita. Caso as funções $g^i(x)$ sejam não-lineares, o seguinte teste oferecido por Arrow e Enthoven pode ser útil para determinar se a qualificação de restrição é satisfeita:

Para um problema de maximização, se

- (a) cada função de restrição $g^i(x)$ for diferenciável e quase-convexa
- (b) existir um ponto x^0 no ortante não negativo tal que todas as restrições são satisfeitas como desigualdades estritas em x^0
- (c) uma das seguintes for verdadeira:
 - (c-i) toda função $g^i(x)$ for convexa
 - (c-ii) as derivadas parciais de cada $g^i(x)$ não forem todas zero quando calculadas em todos os pontos x na região viável

então a qualificação de restrição é satisfeita.

Mais uma vez, esse teste pode ser adaptado com facilidade ao problema de maximização. Para fazer isso, basta trocar a palavra *quase-convexa* por *quase-côncava* na condição (a), e trocar a palavra *convexa* para *côncava* em (c-i).

EXERCÍCIO 13.4

1. Dado o problema:

Minimize	$C = F(x)$	
sujeita a	$G^i(x) \geq r_i$	$(i = 1, 2, \dots, m)$
e	$x > 0$	

 - (a) Converta-o para um problema de maximização.
 - (b) Neste problema, quais são as equivalentes das funções f e g^i no teorema de suficiência de Kuhn-Tucker?
 - (c) Por consequência, quais condições de concavidade-convexidade devem ser impostas às funções F e G^i para fazer com que as condições suficientes para um máximo sejam aplicáveis aqui?
 - (d) Com base no problema acima, como você enunciaria as condições de Kuhn-Tucker suficientes para um *mínimo*?
2. O teorema de suficiência de Kuhn-Tucker aos problemas a seguir?
 - (a) Maximize $\pi = x_1$
 sujeita a $x_1^2 + x_3^2 \leq 1$
 e $x_1, x_2 \geq 0$
 - (b) Minimize $C = (x_1 - 3)^2 + (x_2 - 4)^2$
 sujeita a $x_1 + x_2 \geq 4$
 e $x_1, x_2 \geq 0$

$$\begin{array}{ll}
 \text{(c) Minimize} & C = 2x_1 + x_2 \\
 \text{sujeita a} & x_1^2 - 4x_1 + x_2 \geq 0 \\
 \text{e} & x_1, x_2 \geq 0
 \end{array}$$

3. Quais das seguintes funções são matematicamente aceitáveis como a função objetivo de um problema de *maximização* que se qualifica para a aplicação do teorema de suficiência de Arrow-Enthoven?

(a) $f(x) = x^3 - 2x$

(b) $f(x_1, x_2) = 6x_1 - 9x_2$

(c) $f(x_1, x_2) = x_2 - \ln x_1$ (Nota: Veja Exercício 12.4–4.)

4. A qualificação de restrição de Arrow-Enthoven é satisfeita, dado que as restrições de um problema de *maximização* são:

(a) $x_1^2 + (x_2 - 5)^2 \leq 4$ e $5x_1 + x_2 < 10$

(b) $x_1 + x_2 \leq 8$ e $-x_1x_2 \leq -8$ (Nota: $-x_1x_2$ não é convexa.)

13.5 Funções valor máximo e o teorema do envelope[†]

Uma função valor máximo é uma função objetivo na qual foram atribuídos às variáveis de escolha seus valores ótimos. Esses valores ótimos das variáveis de escolha, por sua vez, são funções das variáveis exógenas e dos parâmetros do problema. Uma vez substituídos os valores ótimos das variáveis de escolha na função objetivo original, esta se torna, indiretamente, uma função apenas dos parâmetros (pela influência dos parâmetros sobre os valores ótimos das variáveis de escolha). Assim, a função valor máximo também é conhecida como a *função objetivo indireta*.

O teorema do envelope para otimização sem restrição

Qual é o significado da função objetivo indireta? Considere que, em qualquer problema de otimização, a função objetivo direta é maximizada (ou minimizada) para um dado conjunto de parâmetros. A função objetivo indireta rastreia todos os valores de máximo da função objetivo à medida que esses parâmetros variam. Por conseguinte, a função objetivo indireta é um “envelope” do conjunto de funções objetivo otimizadas geradas pela variação dos parâmetros do modelo. Para a maioria dos estudantes de economia, a primeira ilustração dessa idéia de um envelope surge da comparação entre curvas de custo de curto e de longo prazo. Normalmente eles aprendem que a curva de custo médio de longo prazo é um envelope de todas as curvas de custo médio de curto prazo (qual parâmetro está variando ao longo do envelope nesse caso?). Uma dedução formal desse conceito é um dos exercícios que faremos nesta seção.

Para ilustrar, considere o seguinte problema de maximização sem restrição com duas variáveis de escolha x e y e um parâmetro ϕ :

$$\text{Maximize } U = f(x, y, \phi) \quad (13.27)$$

A condição necessária de primeira ordem é

$$f_x(x, y, \phi) = f_y(x, y, \phi) = 0 \quad (13.28)$$

Se as condições de segunda ordem forem cumpridas, essas duas equações definem, implicitamente, as soluções

[†] Esta seção do capítulo apresenta uma visão geral do teorema do envelope. Um tratamento mais detalhado desse tópico pode ser encontrado no Capítulo 7 de *The Structure of Economics: A Mathematical Analysis* (3ª edição), de Eugene Silberberg e Wing Suen (McGraw-Hill, 2001), no qual são baseadas partes desta seção.

$$x^* = x^*(\phi) \quad y^* = y^*(\phi) \quad (13.29)$$

Se substituirmos essas soluções na função objetivo, obtemos uma nova função

$$V(\phi) = f(x^*(\phi), y^*(\phi), \phi) \quad (13.30)$$

onde essa função é o valor de f quando os valores de x e y são os que maximizam $f(x, y, \phi)$. Portanto, $V(\phi)$ é a *função valor máximo* (ou função objetivo indireta). Se diferenciarmos V em relação a ϕ , seu único argumento, obtemos

$$\frac{dV}{d\phi} = f_x \frac{\partial x^*}{\partial \phi} + f_y \frac{\partial y^*}{\partial \phi} + f_\phi \quad (13.31)$$

Contudo, sabemos, pelas condições de primeira ordem, que $f_x = f_y = 0$. Portanto, os dois primeiros termos se anulam e o resultado se torna

$$\frac{dV}{d\phi} = f_\phi \quad (13.31')$$

Esse resultado diz que, no ponto ótimo, à medida que ϕ varia, permitindo-se o ajuste de x^* e y^* , a derivada $dV/d\phi$ dá o mesmo resultado que daria se x^* e y^* fossem tratados como constantes. Note que ϕ entra na função valor máximo (13.30) em três lugares: um direto e dois indiretos (por meio de x^* e y^*). A equação (13.31') mostra que, no ponto ótimo, somente o efeito direto de ϕ sobre a função objetivo é que importa. Essa é a essência do teorema do envelope. O teorema do envelope diz que somente os efeitos diretos de uma mudança em uma variável exógena precisam ser considerados, mesmo que a variável exógena também possa entrar na função valor máximo indiretamente como parte da solução das variáveis de escolha endógenas.

A função lucro

Agora vamos aplicar a noção da função valor máximo para obter a função lucro de uma empresa competitiva. Considere o caso em que a empresa utiliza dois insumos: capital, K , e trabalho, L . A função lucro é

$$\pi = Pf(K, L) - wL - rK \quad (13.32)$$

onde P é o preço do produto e w e r são a taxa salarial e a taxa de aluguel, respectivamente.

As condições de primeira ordem são

$$\begin{aligned} \pi_L &= Pf_L(K, L) - w = 0 \\ \pi_K &= Pf_K(K, L) - r = 0 \end{aligned} \quad (13.33)$$

que definem, respectivamente, as equações de demanda de insumos

$$\begin{aligned} L^* &= L^*(w, r, P) \\ K^* &= K^*(w, r, P) \end{aligned} \quad (13.34)$$

Substituindo as soluções K^* e L^* na função objetivo, temos

$$\pi^*(w, r, P) = Pf(K^*, L^*) - wL^* - rK^* \quad (13.35)$$

onde $\pi^*(w, r, P)$ é a *função lucro* (uma função objetivo indireta). A função lucro dá o lucro máximo como uma função das variáveis exógenas w , r e P .

Agora considere o efeito causado por uma variação em w sobre os lucros da empresa. Se diferenciarmos a função lucro original em (13.32) em relação a w , mantendo todas as outras variáveis constantes, obtemos

$$\frac{\partial \pi}{\partial w} = -L \quad (13.36)$$

Contudo, esse resultado não leva em conta a capacidade maximizadora de lucro que a empresa tem, que é substituir capital por trabalho e ajustar o nível de produção de acordo com um comportamento maximizador de lucro.

Ao contrário, visto que $\pi^*(w, r, P)$ é o valor máximo de lucros para quaisquer valores de w, r e P , variações em π^* causadas por uma variação em w levam em conta todas as substituições de capital por trabalho. Para calcular uma variação na função lucro máximo causada por uma variação em w , diferenciamos $\pi^*(w, r, P)$ em relação a w para obter

$$\frac{\partial \pi^*}{\partial w} = (Pf_L - w) \frac{\partial L^*}{\partial w} + (Pf_K - r) \frac{\partial K^*}{\partial w} - L^* \quad (13.37)$$

Pelas condições de primeira ordem (13.33), os dois termos entre parênteses são iguais a zero. Portanto, a equação se torna

$$\frac{\partial \pi^*}{\partial w} = -L^*(w, r, P) \quad (13.38)$$

Esse resultado diz que, na posição que maximiza lucros, uma variação nos lucros relativa a uma variação na taxa salarial é a mesma quer os fatores sejam mantidos constantes, quer seja permitido que variem à medida que o preço do fator varia. Nesse caso, (13.38) mostra que a derivada da função lucro em relação a w é a negativa da função demanda do fator $L^*(w, r, P)$. Seguindo o procedimento precedente, também podemos mostrar os resultados adicionais de estática comparativa:

$$\frac{\partial \pi^*(w, r, P)}{\partial r} = -K^*(w, r, P) \quad (13.39)$$

$$\text{e} \quad \frac{\partial \pi^*(w, r, P)}{\partial P} = f(K^*, L^*) \quad (13.40)$$

As equações (13.38), (13.39) e (13.40) são conhecidas coletivamente como *lema de Hotelling*. Obtivemos essas derivadas de estática comparativa da função lucro permitindo que K^* e L^* se ajustassem a qualquer variação nos parâmetros. Mas é fácil verificar que os mesmos resultados surgirão se diferenciarmos a função lucro (13.35) em relação a cada parâmetro, mantendo, ao mesmo tempo, K^* e L^* constantes. Assim, o lema de Hotelling é, simplesmente, uma outra manifestação do teorema do envelope que encontramos anteriormente em (13.31').

Condição de reciprocidade

Considere novamente nosso problema de maximização sem restrição com duas variáveis.

$$\text{Maximize } U = f(x, y, \phi) \quad [\text{de (13.27)}]$$

onde x e y são as variáveis de escolha e ϕ é um parâmetro. As condições de primeira ordem são $f_x = f_y = 0$, o que implica $x^* = x^*(\phi)$ e $y^* = y^*(\phi)$.

Estamos interessados na estática comparativa referente às direções da variação em $x^*(\phi)$ e $y^*(\phi)$ à medida que ϕ varia e nos efeitos sobre a função valor. A função valor máximo é

$$V(\phi) = f(x^*(\phi), y^*(\phi), \phi) \quad (13.41)$$

Por definição, $V(\phi)$ dá o valor máximo de f para qualquer ϕ dado.

Agora considere uma nova função que retrata a diferença entre o valor real e o valor máximo de U :

$$\Omega(x, y, \phi) = f(x, y, \phi) - V(\phi) \quad (13.42)$$

Essa nova função Ω tem um valor máximo de zero quando $x = x^*$ e $y = y^*$; para quaisquer $x \neq x^*$, $y \neq y^*$, temos $f \leq V$. Nessa estrutura, $\Omega(x, y, \phi)$ pode ser considerada uma função de três variáveis independentes, $x, y \in \phi$. O máximo de $\Omega(x, y, \phi) = f(x, y, \phi) - V(\phi)$ pode ser determinado pelas condições de primeira e de segunda ordem.

As condições de primeira ordem são

$$\Omega_x(x, y, \phi) = f_x = 0 \quad (13.43)$$

$$\Omega_y(x, y, \phi) = f_y = 0$$

$$\text{e} \quad \Omega_\phi(x, y, \phi) = f_\phi - V_\phi = 0 \quad (13.44)$$

Podemos ver que as condições de primeira ordem de nossa nova função Ω em (13.43) nada mais são do que as condições originais de máximo para $f(x, y, \phi)$ em (13.28), ao passo que a condição em (13.44) realmente reafirma o teorema do envelope (13.31'). Essas condições de primeira ordem são válidas sempre que $x = x^*(\phi)$ e $y = y^*(\phi)$. As condições suficientes de segunda ordem são satisfeitas se o hessiano de Ω

$$H = \begin{vmatrix} f_{xx} & f_{xy} & f_{x\phi} \\ f_{yx} & f_{yy} & f_{y\phi} \\ f_{\phi x} & f_{\phi y} & f_{\phi\phi} - V_{\phi\phi} \end{vmatrix}$$

for caracterizado por

$$f_{xx} < 0 \quad f_{xx}f_{yy} - f_{xy}^2 > 0 \quad H < 0$$

Ao obter o hessiano acima, listamos as variáveis na ordem (x, y, ϕ) e, por consequência, o primeiro coeficiente nas condições de segunda ordem, $(\Omega_{xx} =) f_{xx} < 0$ está relacionado à variável x . Se tivéssemos adotado uma ordem de listagem alternativa, então o primeiro-coeficiente poderia ter sido $\Omega_{yy} = f_{yy} < 0$, ou

$$\Omega_{\phi\phi} = f_{\phi\phi} - V_{\phi\phi} < 0 \quad (13.45)$$

Resulta que (13.45) pode nos levar a um resultado que proporciona um modo rápido de chegar a uma conclusão de estática comparativa. Em primeiro lugar, sabemos, por (13.41), que

$$V_\phi(\phi) = f_\phi(x^*(\phi), y^*(\phi), \phi)$$

Diferenciando ambos os lados em relação a ϕ , temos

$$V_{\phi\phi} = f_{\phi x} \frac{\partial x^*}{\partial \phi} + f_{\phi y} \frac{\partial y^*}{\partial \phi} + f_{\phi\phi} \quad (13.46)$$

Usando (13.45) e o teorema de Young, podemos escrever

$$V_{\phi\phi} - f_{\phi\phi} = f_{x\phi} \frac{\partial x^*}{\partial \phi} + f_{y\phi} \frac{\partial y^*}{\partial \phi} > 0 \quad (13.47)$$

Suponha que ϕ entra somente na condição de primeira ordem para x , de modo que $f_{y\phi} = 0$. Então (13.47) se reduz a

$$f_{x\phi} \frac{\partial x^*}{\partial \phi} > 0 \quad (13.48)$$

o que implica que $f_{x\phi}$ e $\partial x^* / \partial \phi$ terão o mesmo sinal. Assim, sempre que virmos o parâmetro ϕ aparecendo somente na condição de primeira ordem relacionada a x e, uma vez determinado o sinal da derivada $f_{x\phi}$ pela função objetivo $U = f(x, y, \phi)$, podemos dizer imediatamente qual é o sinal da derivada de estática comparativa $\partial x^* / \partial \phi$ sem maiores complicações.

Por exemplo, no modelo de maximização de lucro:

$$\pi = Pf(K, L) - wL - rK$$

no qual as condições de primeira ordem são:

$$\pi_L = Pf_L - w = 0$$

$$\pi_K = Pf_K - r = 0$$

a variável exógena w entra somente na condição de primeira ordem $Pf_L - w = 0$, com

$$\frac{\partial \pi_L}{\partial w} = -1$$

Portanto, podemos concluir, por (13.48), que $\partial L^* / \partial w$ também será negativa.

Além disso, se combinarmos o teorema do envelope com o teorema de Young, podemos deduzir uma relação conhecida como a *condição de reciprocidade*: $\partial L^* / \partial r = \partial K^* / \partial w$. Pela função lucro indireta $\pi^*(w, r, P)$, o lema de Hotelling nos dá

$$\pi_w^* = \frac{\partial \pi^*}{\partial w} = -L^*(w, r, P)$$

$$\pi_r^* = \frac{\partial \pi^*}{\partial r} = -K^*(w, r, P)$$

Diferenciando novamente e aplicando o teorema de Young, temos

$$\pi_{wr}^* = -\frac{\partial L^*}{\partial r} = -\frac{\partial K^*}{\partial w} = \pi_{rw}^*$$

ou

$$\frac{\partial L^*}{\partial r} = \frac{\partial K^*}{\partial w} \quad (13.49)$$

Esse resultado é denominado *condição de reciprocidade* porque mostra a simetria entre o efeito cruzado de estática comparativa produzido pelo preço de um único insumo sobre a demanda pelo “outro” insumo. Especificamente, no sentido da estática comparativa, o efeito de r (a taxa de aluguel para o capital K) sobre a demanda ótima para o trabalho L é o mesmo que o efeito de w (a taxa salarial para o trabalho L) sobre a demanda ótima pelo capital K .

O teorema do envelope para otimização com restrição

O teorema do envelope também pode ser obtido para o caso de otimização restrita. Mais uma vez teremos uma função objetivo (U), duas variáveis de escolha (x e y) e um parâmetro (ϕ); exceto que agora introduzimos a seguinte restrição:

$$g(x, y; \phi) = 0$$

O problema torna-se:

$$\begin{aligned} &\text{Maximize } U = f(x, y; \phi) \\ &\text{sujeita a } g(x, y; \phi) = 0 \end{aligned} \quad (13.50)$$

A função de Lagrange para esse problema é

$$Z = f(x, y; \phi) + \lambda[0 - g(x, y; \phi)] \quad (13.51)$$

com condições de primeira ordem

$$\begin{aligned} Z_x &= f_x - \lambda g_x = 0 \\ Z_y &= f_y - \lambda g_y = 0 \\ Z_\lambda &= -g(x, y; \phi) = 0 \end{aligned}$$

Resolvendo esse sistema de equações, temos

$$x = x^*(\phi) \quad y = y^*(\phi) \quad \lambda = \lambda^*(\phi)$$

Substituindo as soluções na função objetivo, obtemos

$$U^* = f(x^*(\phi), y^*(\phi), \phi) = V(\phi) \quad (13.52)$$

onde $V(\phi)$ é a função objetivo indireta, uma função valor máximo. Esse é o valor máximo de y para qualquer ϕ e x_i que satisfaçam a restrição.

Como $V(\phi)$ varia à medida que ϕ varia? Primeiro, diferenciamos V em relação a ϕ :

$$\frac{dV}{d\phi} = f_x \frac{\partial x^*}{\partial \phi} + f_y \frac{\partial y^*}{\partial \phi} + f_\phi \quad (13.53)$$

Contudo, nesse caso, (13.53) não será simplificada para $dV/d\phi = f_\phi$, visto que, na otimização restrita, não é necessário que $f_x = f_y = 0$ (veja Tabela 12.1). Mas, se substituirmos as soluções para x e y na restrição (produzindo uma identidade), obteremos

$$g(x^*(\phi), y^*(\phi), \phi) \equiv 0$$

e, diferenciando-a em relação a ϕ , temos

$$g_x \frac{\partial x^*}{\partial \phi} + g_y \frac{\partial y^*}{\partial \phi} + g_\phi \equiv 0 \quad (13.54)$$

Se multiplicarmos (13.54) por λ , combinarmos o resultado com (13.53) e rearranjarmos os termos, obteremos

$$\frac{dV}{d\phi} = (f_x - \lambda g_x) \frac{\partial x^*}{\partial \phi} + (f_y - \lambda g_y) \frac{\partial y^*}{\partial \phi} + f_\phi - \lambda g_\phi = Z_\phi \quad (13.55)$$

onde Z_ϕ é a derivada parcial da função de Lagrange em relação a ϕ , mantendo constantes todas as outras variáveis. Esse resultado tem as mesmas características de (13.31) e, em virtude das condições de primeira ordem, ele se reduz a

$$\frac{dV}{d\phi} = Z_\phi \quad (13.55')$$

que representa o teorema do envelope na estrutura da otimização restrita. Contudo, note que, no presente caso, a função de Lagrange substitui a função objetivo no cálculo da função objetivo indireta.

Conquanto os resultados em (13.55) guardem um bom paralelo com o caso sem restrição, é importante notar que alguns dos resultados de estática comparativa dependem criticamente de os parâmetros aparecerem somente na função objetivo, ou somente nas restrições, ou em ambas. Se um parâmetro aparecer somente na função objetivo, então os resultados de estática comparativa são os mesmos do caso sem restrição. Contudo, se o parâmetro aparecer na restrição, a relação

$$V_{\phi\phi} \geq f_{\phi\phi}$$

não valerá mais.

Interpretação do multiplicador de Lagrange

No problema de escolha do consumidor no Capítulo 12, obtivemos o resultado de que o multiplicador de Lagrange λ representava a variação no valor da função de Lagrange quando o orçamento do consumidor mudava. Interpretamos λ como a utilidade marginal de insumo. Agora vamos deduzir uma interpretação mais geral do multiplicador de Lagrange com o auxílio do teorema do envelope. Considere o problema

$$\begin{aligned} &\text{Maximize } U = f(x, y) \\ &\text{sujeita a } g(x, y) = c \end{aligned}$$

onde c é uma constante. A função de Lagrange para esse problema é

$$Z = f(x, y) + \lambda[c - g(x, y)] \quad (13.56)$$

As condições de primeira ordem são

$$\begin{aligned} Z_x &= f_x(x, y) - \lambda g_x(x, y) = 0 \\ Z_y &= f_y(x, y) - \lambda g_y(x, y) = 0 \\ Z_\lambda &= c - g(x, y) = 0 \end{aligned} \quad (13.57)$$

Das duas primeiras equações em (13.57), obtemos

$$\lambda = \frac{f_x}{g_x} = \frac{f_y}{g_y} \quad (13.58)$$

que nos dá uma condição: a inclinação da curva de nível (curva de indiferença) da função objetivo deve ser igual à inclinação da restrição no ponto ótimo.

As equações (13.57) definem implicitamente as soluções

$$x^* = x^*(c) \quad y^* = y^*(c) \quad \lambda^* = \lambda^*(c) \quad (13.59)$$

Substituindo (13.59) na função de Lagrange, temos a função valor máximo,

$$V(c) = Z^*(c) = f(x^*(c), y^*(c)) + \lambda^*(c) [c - g(x^*(c), y^*(c))] \quad (13.60)$$

Diferenciando em relação a c , temos como resultado

$$\begin{aligned} \frac{dV}{dc} = \frac{dZ^*}{dc} = f_x \frac{\partial x^*}{\partial c} + f_y \frac{\partial y^*}{\partial c} + [c - g(x^*(c), y^*(c))] \frac{\partial \lambda^*}{\partial c} \\ - \lambda^*(c) g_x \frac{\partial x^*}{\partial c} - \lambda^*(c) g_y \frac{\partial y^*}{\partial c} + \lambda^*(c) \frac{dc}{dc} \end{aligned}$$

Rearranjando, obtemos

$$\frac{dZ^*}{dc} = [f_x - \lambda^* g_x] \frac{\partial x^*}{\partial c} + [f_y - \lambda^* g_y] \frac{\partial y^*}{\partial c} + [c - g(x^*, y^*)] \frac{\partial \lambda^*}{\partial c} + \lambda^*$$

Por (13.57), os três termos entre colchetes são todos iguais a zero. Portanto, nossa expressão é simplificada para

$$\frac{dV}{dc} = \frac{dZ^*}{dc} = \lambda^* \quad (13.61)$$

que mostra que o valor ótimo λ^* mede a taxa de variação do valor máximo da função objetivo quando c varia, e é por essa razão que ele é denominado o “preço sombra” de c . Note que, nesse caso, c entra no problema somente através da restrição; não é um argumento da função objetivo original.

13.6 Dualidade e o teorema do envelope

A função dispêndio de um consumidor e sua função utilidade indireta exemplificam as funções valor mínimo e valor máximo para *problemas duais*.[†] Uma função dispêndio especifica o dispêndio mínimo requerido para obter um nível fixo de utilidade, dada a função utilidade e os preços de bens de consumo. Uma função utilidade indireta especifica a utilidade máxima que pode ser obtida dados preços, renda e a função utilidade.

O problema primordial

Seja $U(x, y)$ uma função utilidade na qual x e y são bens de consumo. O consumidor tem um orçamento B e enfrenta preços de mercado P_x e P_y para os bens x e y , respectivamente. Esse problema será considerado o *problema primordial*:

$$\begin{aligned} &\text{Maximize } U = U(x, y) \\ &\text{sujeita a } P_x x + P_y y = B \quad [\text{Primordial}] \end{aligned} \quad (13.62)$$

Para esse problema, temos a função de Lagrange conhecida

$$Z = U(x, y) + \lambda(B - P_x x - P_y y)$$

[†] Dualidade em teoria econômica é a relação entre dois problemas de otimização restrita. Se um dos problemas exigir maximização restrita, o outro problema exigirá minimização restrita. A estrutura e a solução de qualquer dos problemas podem proporcionar informações sobre a estrutura e a solução do outro problema.

As condições de primeira ordem são

$$\begin{aligned} Z_x &= U_x - \lambda P_x = 0 \\ Z_y &= U_y - \lambda P_y = 0 \\ Z_\lambda &= B - P_x x - P_y y = 0 \end{aligned} \quad (13.63)$$

Esse sistema de equações define implicitamente uma solução para x^m , y^m e λ^m como uma função das variáveis exógenas B , P_x , P_y :

$$\begin{aligned} x^m &= x^m(P_x, P_y, B) \\ y^m &= y^m(P_x, P_y, B) \\ \lambda^m &= \lambda^m(P_x, P_y, B) \end{aligned}$$

As soluções x^m e y^m são as funções demanda comuns de consumidor, denominadas, às vezes, funções demanda “marshallianas”, daí o índice m .

Substituindo as soluções x^m e y^m na função utilidade, temos

$$U^* = U^*(x^m(P_x, P_y, B), y^m(P_x, P_y, B)) \equiv V(P_x, P_y, B) \quad (13.64)$$

onde V é a função utilidade indireta – uma função valor máximo que mostra a utilidade máxima que se pode obter no problema (13.62). Voltaremos a essa função mais tarde.

O problema dual

Agora considere um *problema dual* relacionado, ou seja, o de um consumidor cujo objetivo é minimizar o dispêndio em x e y mantendo, ao mesmo tempo, um nível fixo de utilidade U^* , obtido de (13.64) do problema primordial:

$$\begin{aligned} \text{Minimize } E &= P_x x + P_y y \\ \text{sujeita a } U(x, y) &= U^* \quad [\text{Dual}] \end{aligned} \quad (13.65)$$

Sua função de Lagrange é

$$Z^d = P_x x + P_y y + \mu [U^* - U(x, y)]$$

e as condições de primeira ordem são

$$\begin{aligned} Z_x^d &= P_x - \mu U_x = 0 \\ Z_y^d &= P_y - \mu U_y = 0 \\ Z_\mu^d &= U^* - U(x, y) = 0 \end{aligned} \quad (13.66)$$

Esse sistema de equações define implicitamente um conjunto de valores de solução representados por x^h , y^h e λ^h :

$$\begin{aligned} x^h &= x^h(P_x, P_y, U^*) \\ y^h &= y^h(P_x, P_y, U^*) \\ \mu^h &= \mu^h(P_x, P_y, U^*) \end{aligned}$$

Aqui, x^h e y^h são as funções demanda compensadas (“renda real” mantida constante). Elas são comumente denominadas funções demanda “hicksianas”, daí o índice h .

Substituindo x^h e y^h na função objetivo do problema dual, temos

$$P_x x^h(P_x, P_y, U^*) + P_y y^h(P_x, P_y, U^*) \equiv E(P_x, P_y, U^*) \quad (13.67)$$

onde E é a função dispêndio – uma função valor mínimo que mostra o dispêndio mínimo necessário para obter o nível de utilidade U^* .

Dualidade

Se tomarmos as duas primeiras equações em (13.63) e em (13.64) e eliminarmos os multiplicadores de Lagrange podemos, escrever

$$\frac{P_x}{P_y} = \frac{U_x}{U_y} \quad (13.68)$$

Essa é a condição de tangência na qual o consumidor escolhe o conjunto ótimo no qual a inclinação da curva de indiferença é idêntica à inclinação da restrição orçamentária. A condição de tangência é idêntica para ambos os problemas. Assim, quando o nível de utilidade visado no problema de minimização é igualado aos valores U^* obtidos do problema de maximização, obtemos

$$\begin{aligned} x^m(P_x, P_y, B) &= x^h(P_x, P_y, U^*) \\ y^m(P_x, P_y, B) &= y^h(P_x, P_y, U^*) \end{aligned} \quad (13.69)$$

isto é, as soluções para o problema de maximização, bem como para o problema de minimização, produzem valores idênticos para x e y . Contudo, as soluções são funções de variáveis exógenas diferentes, portanto exercícios de estática comparativa geralmente produzirão resultados diferentes.

O fato de os valores de solução para x e y no problema primordial e no problema dual serem determinados pelo ponto de tangência das mesmas curva de indiferença e linha de restrição orçamentária significa que o dispêndio minimizado no problema dual é igual ao orçamento B dado do problema primordial:

$$E(P_x, P_y, U^*) = B \quad (13.70)$$

Esse resultado é paralelo ao resultado em (13.64), o qual revela que o valor maximizado da utilidade V no problema primordial é igual ao nível desejado dado da utilidade U^* no problema dual.

Conquanto os valores de solução de x e y sejam idênticos nos dois problemas, não se pode dizer o mesmo sobre os multiplicadores de Lagrange. Pela primeira equação em (13.63) e em (13.66), podemos calcular $\lambda = U_x/P_x$, mas $\mu = P_x/U_x$. Assim, os valores de solução de λ e μ são mutuamente recíprocos:

$$\lambda = \frac{1}{\mu} \quad \text{ou} \quad \lambda^m = \frac{1}{\mu^h} \quad (13.71)$$

Identidade de Roy

Uma aplicação do teorema do envelope é a dedução da identidade de Roy. A identidade de Roy afirma que a função demanda marshalliana do consumidor individual é igual à negativa da razão entre as duas derivadas parciais da função valor máximo.

Substituindo os valores ótimos x^m , y^m e λ^m na função de Lagrange de (13.62), temos

$$V(P_x, P_y, B) = U(x^m, y^m) + \lambda^m(B - P_x x^m - P_y y^m) \quad (13.72)$$

Quando diferenciamos (13.72) em relação a P_x , encontramos

$$\begin{aligned} \frac{\partial V}{\partial P_x} &= (U_x - \lambda^m P_x) \frac{\partial x^m}{\partial P_x} (U_y - \lambda^m P_y) \frac{\partial y^m}{\partial P_x} \\ &\quad + (B - P_x x^m - P_y y^m) \frac{\partial \lambda^m}{\partial P_x} - \lambda^m x^m \end{aligned}$$

No ponto ótimo, as condições de primeira ordem (13.63) nos permitem simplificar essa expressão para

$$\frac{\partial V}{\partial P_x} = -\lambda^m x^m$$

Em seguida, diferenciamos a função valor em relação a B para obter

$$\begin{aligned} \frac{\partial V}{\partial B} &= (U_x - \lambda^m P_x) \frac{\partial x^m}{\partial B} (U_y - \lambda^m P_y) \frac{\partial y^m}{\partial B} \\ &\quad + (B - P_x x^m - P_y y^m) \frac{\partial \lambda^m}{\partial B} + \lambda^m \end{aligned}$$

Novamente, no ponto ótimo, (13.63) nos permite simplificar essa expressão para

$$\frac{\partial V}{\partial B} = \lambda^m$$

Tomando a razão entre essas duas derivadas parciais, constatamos que

$$\frac{\partial V / \partial P_x}{\partial V / \partial B} = -x^m \quad (13.73)$$

Esse resultado, conhecido como *identidade de Roy*, mostra que a demanda marshalliana pela mercadoria x é a negativa da razão entre duas derivadas parciais da função valor máximo V em relação a P_x e B , respectivamente. Em vista da simetria entre x e y no problema, também pode ser escrito um resultado análogo a (13.73) para y^m , a função demanda marshalliana para y . É claro que poderíamos ter chegado a esse resultado diretamente aplicando o teorema do envelope.

O lema de Shephard

Na Seção 13.5, obtivemos o lema de Hotelling, que afirma que as derivadas parciais do valor máximo da função lucro dão como resultado as funções insumo-demanda da empresa e as funções oferta. Uma abordagem semelhante aplicada à função dispêndio resulta no lema de Shephard.

Considere o problema de minimização do consumidor (13.65). A função de Lagrange é

$$Z^d = P_x x + P_y y + \mu [U^* - U(x, y)]$$

Pelas condições de primeira ordem, as seguintes soluções estão definidas implicitamente

$$\begin{aligned} x^h &= x^h(P_x, P_y, U^*) \\ y^h &= y^h(P_x, P_y, U^*) \\ \mu^h &= \mu^h(P_x, P_y, U^*) \end{aligned}$$

Substituir essas soluções na função de Lagrange dá como resultado a função dispêndio:

$$E(P_x, P_y, U^*) = P_x x^h + P_y y^h + \mu^h [U^* - U(x^h, y^h)]$$

Tomando as derivadas parciais dessa função em relação a P_x e P_y e calculando-as no ponto ótimo, constatamos que $\partial E/\partial P_x$ e $\partial E/\partial P_y$ representam as demandas hicksianas do consumidor:

$$\begin{aligned}\frac{\partial E}{\partial P_x} &= (P_x - \mu^h U_x) \frac{\partial x^h}{\partial P_x} + (P_y - \mu^h U_y) \frac{\partial y^h}{\partial P_x} [U^* - U(x^h, y^h)] \frac{\partial \mu^h}{\partial P_x} + x^h \\ &= (0) \frac{\partial x^h}{\partial P_x} + (0) \frac{\partial y^h}{\partial P_x} + (0) \frac{\partial \mu^h}{\partial P_x} + x^h = x^h\end{aligned}\quad (13.74)$$

e

$$\begin{aligned}\frac{\partial E}{\partial P_y} &= (P_x - \mu^h U_x) \frac{\partial x^h}{\partial P_y} + (P_y - \mu^h U_y) \frac{\partial y^h}{\partial P_y} [U^* - U(x^h, y^h)] \frac{\partial \mu^h}{\partial P_y} + y^h \\ &= (0) \frac{\partial x^h}{\partial P_y} + (0) \frac{\partial y^h}{\partial P_y} + (0) \frac{\partial \mu^h}{\partial P_y} + y^h = y^h\end{aligned}\quad (13.74')$$

Por fim, diferenciar E em relação à restrição U^* resulta em μ^h , o custo marginal da restrição

$$\begin{aligned}\frac{\partial E}{\partial U^*} &= (P_x - \mu^h U_x) \frac{\partial x^h}{\partial U^*} + (P_y - \mu^h U_y) \frac{\partial y^h}{\partial U^*} \\ &\quad + [U^* - U(x^h, y^h)] \frac{\partial \mu^h}{\partial U^*} + \mu^h \\ &= (0) \frac{\partial x^h}{\partial U^*} + (0) \frac{\partial y^h}{\partial U^*} + (0) \frac{\partial \mu^h}{\partial U^*} + \mu^h = \mu^h\end{aligned}\quad (13.74'')$$

Juntas, as três derivadas parciais (13.74), (13.74') e (13.74'') são denominadas lema de Shephard.

EXEMPLO 1

Considere um consumidor cuja função utilidade é $U = xy$ e que enfrenta uma restrição orçamentária de B e os preços dados P_x e P_y .

O problema da escolha é

$$\begin{array}{ll}\text{Maximize} & U = xy \\ \text{sujeita a} & P_x x + P_y y = B\end{array}$$

A função de Lagrange para esse problema é

$$Z = xy + \lambda(B - P_x x - P_y y)$$

As condições de primeira ordem são

$$\begin{aligned}Z_x &= y - \lambda P_x = 0 \\ Z_y &= x - \lambda P_y = 0 \\ Z_\lambda &= B - P_x x - P_y y = 0\end{aligned}$$

Resolver as condições de primeira ordem resulta nas seguintes soluções:

$$x^m = \frac{B}{2P_x} \quad y^m = \frac{B}{2P_y} \quad \lambda^m = \frac{B}{2P_x P_y}$$

onde x^m e y^m são as funções demanda marshalliana do consumidor. Para a condição de segunda ordem, visto que o hessiano aumentado é

$$|\bar{H}| = \begin{vmatrix} 0 & 1 & -P_x \\ 1 & 0 & -P_y \\ -P_x & -P_y & 0 \end{vmatrix} = 2P_x P_y > 0$$

a solução realmente representa um máximo.[†]

Agora podemos obter a função utilidade indireta para esse problema substituindo x^m e y^m na função utilidade:

$$V(P_x, P_y, B) = \left(\frac{B}{2P_x}\right) \left(\frac{B}{2P_y}\right) = \frac{B^2}{4P_x P_y} \quad (13.75)$$

onde V denota a utilidade maximizada. Uma vez que V representa a utilidade maximizada, podemos estabelecer $V = U^*$ em (13.75) para obter $B^2/4P_x P_y = U^*$, e então rearranjar os termos para expressar B como

$$B = (4P_x P_y U^*)^{1/2} = 2P_x^{1/2} P_y^{1/2} U^{*1/2}$$

Agora, pense no problema dual de minimização do dispêndio do consumidor. No problema dual, a função dispêndio mínimo E deve ser igual à quantia equivalente ao orçamento dado B do problema primordial. Por conseguinte, pela equação precedente, podemos concluir imediatamente que

$$E(P_x, P_y, U^*) = B = 2P_x^{1/2} P_y^{1/2} U^{*1/2} \quad (13.76)$$

Agora vamos usar esse exemplo para verificar a identidade de Roy (13.73)

$$x^m = - \frac{\partial V / \partial P_x}{\partial V / \partial B}$$

Tomando as derivadas parciais relevantes de V , encontramos

$$\frac{\partial V}{\partial P_x} = - \frac{B^2}{4P_x^2 P_y}$$

e

$$\frac{\partial V}{\partial B} = \frac{B}{2P_x P_y}$$

[†] Note que, aqui (e no Exemplo 2 na página a seguir), as novas linhas e colunas do hessiano aparecem na terceira linha e na terceira coluna, em vez de na primeira linha e primeira coluna, como em (12.19). Isso é o resultado de considerar o multiplicador de Lagrange como a última variável, e não como a primeira, como fizemos nos capítulos anteriores. O Exercício 12.3–3 mostra que as duas expressões alternativas para o hessiano aumentado podem ser transformadas uma na outra por operações elementares de linha, sem afetar seu valor. Todavia, quando aparecem mais de duas variáveis de escolha em um problema, é preferível usar o formato (12.19) porque isso facilita escrever os menores principais líderes aumentados.

A negativa da razão entre essas duas parciais é

$$-\frac{\frac{\partial V}{\partial P_x}}{\frac{\partial V}{\partial B}} = -\frac{\left(\frac{B^2}{4P_x^2 P_y}\right)}{\left(\frac{B}{2P_x P_y}\right)} = \frac{B}{2P_x} = x^m$$

Assim, constatamos que a identidade de Roy realmente é válida.

EXEMPLO 2

Agora considere o problema dual de minimização de custo dado um nível fixo de utilidade relacionado ao Exemplo 1. Denotando o nível visado de utilidade por U^* , o problema é:

$$\begin{array}{ll} \text{Minimize} & P_x x + P_y y \\ \text{sujeita a} & xy = U^* \end{array}$$

A função de Lagrange para o problema é

$$Z^d = P_x x + P_y y + \mu(U^* - xy)$$

As condições de primeira ordem são

$$Z_x^d = P_x - \mu y = 0$$

$$Z_y^d = P_y - \mu x = 0$$

$$Z_\mu^d = U^* - xy = 0$$

Resolvendo o sistema equações para x , y e μ , obtemos

$$\begin{aligned} x^h &= \left(\frac{P_y U^*}{P_x}\right)^{\frac{1}{2}} \\ y^h &= \left(\frac{P_x U^*}{P_y}\right)^{\frac{1}{2}} \\ \mu^h &= \left(\frac{P_x P_y}{U^*}\right)^{\frac{1}{2}} \end{aligned} \quad (13.77)$$

onde x^h e y^h são as funções demanda compensadas (Hicksianas) do consumidor. Verificando a condição de segunda ordem para um mínimo, encontramos

$$|\bar{H}| = \begin{vmatrix} 0 & -\mu & -y \\ -\mu & 0 & -x \\ -y & -x & 0 \end{vmatrix} = -2xy\mu < 0$$

Assim, a condição suficiente para um mínimo é satisfeita.

Substituindo x^h e y^h na função objetivo original, temos a função valor mínimo, ou função dispêndio

$$\begin{aligned}
 E &= P_x x^h + P_y y^h = P_x \left(\frac{P_y U^*}{P_x} \right)^{1/2} + P_y \left(\frac{P_x U^*}{P_y} \right)^{1/2} \\
 &= (P_x P_y U^*)^{1/2} + (P_x P_y U^*)^{1/2} \\
 &= 2 P_x^{1/2} P_y^{1/2} U^{*1/2}
 \end{aligned} \tag{13.76'}$$

Note que esse resultado é idêntico a (13.76) no Exemplo 1. A única diferença está no processo utilizado para obter o resultado. A equação (13.76') é obtida diretamente de um problema de minimização de dispêndio, ao passo que (13.76) é deduzida indiretamente, via a relação dual, de um problema de maximização de utilidade.

Agora usaremos este exemplo para testar a validade do lema de Shephard (13.74), (13.74'), e (13.74''). Diferenciando a função dispêndio em (13.76') em relação a P_x , P_y e U^* , respectivamente, e relacionando as derivadas parciais resultantes com (13.77), encontramos

$$\begin{aligned}
 \frac{\partial E(P_x, P_y, U^*)}{\partial P_x} &= \frac{P_y^{1/2} U^{*1/2}}{P_x^{1/2}} = x^h \\
 \frac{\partial E(P_x, P_y, U^*)}{\partial P_y} &= \frac{P_x^{1/2} U^{*1/2}}{P_y^{1/2}} = y^h \\
 \frac{\partial E(P_x, P_y, U^*)}{\partial U^*} &= \frac{P_x^{1/2} P_y^{1/2}}{U^{*1/2}} = \mu^h
 \end{aligned}$$

Assim, o lema de Shephard é válido neste exemplo.

EXERCÍCIO 13.6

1. Um consumidor tem a seguinte função utilidade: $U(x, y) = x(y + 1)$, onde x e y são quantidades de dois bens de consumo cujos preços são P_x e P_y , respectivamente. O consumidor também tem um orçamento de B . Por conseguinte, a função de Lagrange do consumidor é

$$x(y + 1) + \lambda(B - P_x x - P_y y)$$

- (a) Encontre pelas condições de primeira ordem, expressões para as funções demanda. Que tipo de bem é y ? Em particular, o que acontece quando $P_y > B$?

- (b) Verifique se esse é um máximo verificando as condições de segunda ordem. Substituindo x^* e y^* na função utilidade, encontre uma expressão para a função utilidade indireta

$$U^* = U(P_x, P_y, B)$$

e deduza uma expressão para a função dispêndio

$$E = E(P_x, P_y, U^*)$$

- (c) Esse problema pode ser reformulado como o seguinte problema dual

$$\text{Minimize} \quad P_x x + P_y y$$

$$\text{sujeita a} \quad x(y + 1) = U^*$$

Encontre os valores de x e y que resolvem esse problema de minimização e mostre que os valores de x e y são iguais às derivadas parciais da função dispêndio, $\partial E / \partial P_x$ e $\partial E / \partial P_y$, respectivamente.

13.7 Algumas observações finais

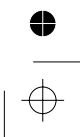
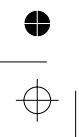
Nesta parte do livro, abordamos as técnicas básicas de otimização. A jornada, até certo ponto árdua, nos conduziu (1) do caso de uma única variável de escolha ao caso mais geral de n variáveis; (2) da função objetivo polinomial às exponenciais e logarítmicas; e (3) da variedade de extremo sem restrições até a de extremo restrito.

Grande parte dessa discussão girou em torno dos métodos “clássicos” de otimização que têm o cálculo diferencial como pilar de sustentação e as derivadas de várias ordens como ferramentas primárias. Uma desvantagem dessa abordagem da otimização é sua natureza essencialmente míope. Conquanto as condições de primeira e segunda ordem, em termos de derivadas ou diferenciais, normalmente possam localizar extremos relativos ou locais sem dificuldade, quase sempre é preciso mais informações ou fazer mais investigações para identificar extremos absolutos ou globais. A discussão detalhada de concavidade, convexidade, quase-concavidade e quase-convexidade tem o propósito de servir como base sólida para partir do reino dos extremos relativos para o dos absolutos.

Uma limitação mais séria da abordagem do cálculo é sua incapacidade de enfrentar restrições na forma de desigualdade. Por essa razão, o enunciado da restrição orçamentária no modelo de maximização de utilidade, por exemplo, considera que o dispêndio total é exatamente *igual* (e não “menor ou igual”) a uma quantia especificada. Em outras palavras, a limitação da abordagem do cálculo faz com que seja necessário negar ao consumidor a opção de poupar parte de seus fundos disponíveis. E, pela mesma razão, a abordagem clássica não nos permite especificar explicitamente que as variáveis de escolha devem ser não-negativas, como é adequado em grande parte das análises econômicas.

Felizmente nos libertamos dessas limitações quando introduzimos a moderna técnica de otimização conhecida como programação não-linear. Com ela podemos admitir abertamente restrições de desigualdade no problema, incluindo restrições de não-negatividade às variáveis de escolha. É óbvio que isso representa um passo gigantesco na direção do desenvolvimento da metodologia de otimização.

Ainda assim, mesmo com a programação não-linear, a estrutura analítica continua estática. O problema e sua solução referem-se somente ao estado ótimo em um único ponto do tempo e não pode abordar a questão de como um agente otimizador deveria se comportar, sob determinadas circunstâncias, durante um período de tempo. Esta última questão pertence ao reino da *otimização dinâmica*, que não podemos utilizar antes de aprender o básico da análise dinâmica – a análise dos movimentos das variáveis ao longo do tempo. De fato, à parte sua aplicação à otimização dinâmica, a análise dinâmica é, em si, um importante ramo da análise econômica. Por essa razão, voltaremos nossa atenção ao tópico da análise dinâmica agora, na Parte 5.



Análise dinâmica



CAPÍTULO 14

Economia dinâmica e cálculo integral

O termo *dinâmica*, como aplicado à análise econômica, teve significados diferentes em épocas diferentes e para diferentes economistas.[†] Contudo, hoje o termo é utilizado comumente com referência ao tipo de análise no qual o objetivo é ou traçar e estudar as trajetórias temporais específicas das variáveis ou determinar se, dado tempo suficiente, essas variáveis tenderão a convergir para certos valores (de equilíbrio). Esse tipo de informação é importante porque preenche uma grande lacuna que prejudicava nosso estudo de estática e de estática comparativa. Nesta última, sempre adotamos a premissa arbitrária de que o processo de ajuste econômico leva inevitavelmente a um equilíbrio. Em uma análise dinâmica, a questão da “atingibilidade” tem de ser encarada de frente em vez de apenas admitida.

Um aspecto notável da análise dinâmica é a *datação* das variáveis, que põe em cena a consideração explícita do *tempo*. Isso pode ser feito de duas maneiras: o tempo pode ser considerado como uma variável *contínua* ou como uma variável *discreta*. No primeiro caso, alguma coisa está acontecendo com a variável em cada *ponto* do tempo (tal como no caso de juro composto contínuo); ao passo que, no último, a variável sofre uma variação apenas uma vez dentro de um *período* de tempo (por exemplo, o juro é adicionado somente ao final de cada seis meses). Um desses conceitos de tempo pode ser mais adequado que o outro em certos contextos.

Em primeiro lugar, discutiremos o caso do tempo contínuo, ao qual são pertinentes as técnicas matemáticas do *cálculo integral* e das *equações diferenciais*. Mais adiante, nos Capítulos 17 e 18, estudaremos o caso do tempo discreto, que utiliza os métodos de *equações de diferenças*.

14.1 Dinâmica e integração

Em um modelo estático, o problema, em termos gerais, é encontrar os valores das variáveis endógenas que satisfaçam alguma (ou algumas) condição de equilíbrio especificada. Aplicado ao contexto de modelos de otimização, a tarefa passa a ser encontrar os valores das variáveis de escolha que maximizem (ou minimizem) uma função objetivo específica – considerando a condição de primeira ordem como a condição de equilíbrio. Em um modelo dinâmico, por outro lado, o problema envolve, por sua vez, o esboço da trajetória temporal de alguma variável, tendo como base um padrão de variação conhecido (digamos, uma determinada taxa de variação instantânea).

Vamos esclarecer com um exemplo. Suponha que saibamos que o tamanho da população H varia ao longo do tempo à taxa

[†] Machlup, Fritz. “Statics and Dynamics: Kaleidoscopic Words”. *Southern Economic Journal*, p. 91–110, out. 1959; impresso novamente em Machlup. *Essays on Economic Semantics*. Englewood Cliffs, N.J., Prentice-Hall, Inc.: 1963, p. 9–42.

$$\frac{dH}{dt} = t^{-1/2} \quad (14.1)$$

Então tentamos encontrar qual, ou quais, trajetória temporal da população $H = H(t)$ pode resultar na taxa de variação em (14.1).

Você perceberá que, se, em primeiro lugar, conhecermos a função $H = H(t)$, a derivada dH/dt pode ser encontrada por diferenciação. Mas, no problema que enfrentamos agora, a situação se inverteu completamente: devemos descobrir a função *primitiva* a partir de uma função *derivada* dada, e não o contrário. Em termos matemáticos, agora precisamos do exato oposto do método de diferenciação, ou do cálculo diferencial.

O método relevante, conhecido como *integração* ou *cálculo integral*, será estudado neste capítulo. Por enquanto, vamos nos contentar com a observação de que a função $H(t) = 2t^{1/2}$ realmente tem uma derivada da forma apresentada em (14.1) e, portanto, aparentemente se qualifica como uma solução para o nosso problema. A dificuldade é que também existem funções similares, tais como $H(t) = 2t^{1/2} + 15$ ou $H(t) = 2t^{1/2} + 99$ ou, de modo mais geral,

$$H(t) = 2t^{1/2} + c \quad (c = \text{uma constante arbitrária}) \quad (14.2)$$

que possuem exatamente a mesma derivada = (14.1). Por conseguinte, não é possível determinar nenhuma trajetória temporal única, a menos que, de algum modo, o valor da constante c possa se tornar definido. Para tal, é preciso introduzir informações adicionais no modelo, usualmente sob a forma do que denominamos uma *condição inicial* ou *condição de fronteira*.

Se soubermos qual é a população inicial $H(0)$ – isto é, o valor de H em $t = 0$, digamos, $H(0) = 100$ – então podemos determinar o valor da constante c . Estabelecendo $t = 0$ em (14.2), temos

$$H(0) = 2(0)^{1/2} + c = c$$

Mas, se $H(0) = 100$, então $c = 100$, e (14.2) torna-se

$$H(t) = 2t^{1/2} + 100 \quad (14.2')$$

onde a constante já não é mais arbitrária. Em termos mais gerais, para qualquer população inicial dada, $H(0)$, a trajetória temporal será

$$H(t) = 2t^{1/2} + H(0) \quad (14.2'')$$

Assim, neste exemplo, o tamanho da população H em qualquer ponto do tempo consistirá na soma da população inicial $H(0)$ e de um outro termo que envolve a variável de tempo t . Tal trajetória temporal realmente traça o itinerário completo da variável H ao longo do tempo e, assim, ela verdadeiramente representa a solução para nosso modelo dinâmico. [A equação (14.1) também é uma função de t . Por que *ela* não pode ser considerada também uma solução?]

Embora simples, esse exemplo da população ilustra a quintessência dos problemas de dinâmica econômica. Dado o padrão de comportamento de uma variável ao longo do tempo, procuramos encontrar uma função que descreva a trajetória temporal da variável. Durante o processo, encontraremos uma ou mais constantes arbitrárias, mas, se dispusermos de informações adicionais suficientes sob a forma de *condições iniciais*, será possível definir essas constantes arbitrárias.

Nos tipos mais simples de problema, tal como o que acabamos de citar, a solução pode ser encontrada pelo método do cálculo integral, que trata do processo de pesquisar o caminho de uma dada função derivada até sua função primitiva. Em casos mais complicados, também podemos recorrer a técnicas conhecidas do ramo da matemática estreitamente relacionado conhecido como *equações diferenciais*. Visto que uma equação diferencial é definida como qualquer equação que contenha expressões diferenciais ou derivadas, (14.1) com certeza se qualifica ao título; por consequência, encontrando sua solução, já resolvemos, de fato, uma equação diferencial, se bem que extremamente simples.

Agora vamos prosseguir com o estudo dos conceitos básicos do cálculo integral. Uma vez que já discutimos cálculo diferencial tendo x (e não t) como a variável independente, por questão de simetria, aqui também usaremos x . Contudo, por conveniência, nesta discussão denotaremos as funções primitiva e derivada por $F(x)$ e $f(x)$, respectivamente, em vez de distingui-las pela utilização do símbolo de “linha” ($'$).

14.2 Integrais indefinidas

A natureza das integrais

Já mencionamos que a integração é o inverso da diferenciação. Se a diferenciação de uma dada função primitiva $F(x)$ resultar na derivada $f(x)$, podemos “integrar” $f(x)$ para encontrar $F(x)$, contanto que disponhamos de informações adequadas para definir a constante arbitrária que surgirá no processo de integração. A função $F(x)$ é denominada uma *integral* (ou *antiderivada*) da função $f(x)$. Assim, esses dois tipos de processo podem ser equiparados a dois modos de estudar uma árvore genealógica: a *integração* envolve traçar a ascendência (os antepassados) da função $f(x)$, enquanto a *diferenciação* busca a descendência da função $F(x)$. Mas note essa diferença – enquanto a função primitiva $F(x)$ (diferenciável) invariavelmente produz um só descendente, a saber, uma única derivada $f(x)$, a função derivada $f(x)$ pode ser traçada até um número infinito de antecedentes possíveis por meio da integração porque, se $F(x)$ for uma integral de $f(x)$, então $F(x)$ mais qualquer constante também deve ser, como vimos em (14.2).

Precisamos de uma notação especial para denotar a integração requerida de $f(x)$ em relação a x . A notação padrão é

$$\int f(x) \, dx$$

O símbolo à esquerda – um S alongado (com a conotação de soma, a ser explicada mais adiante) – é denominado o *signal de integral*, ao passo que a parte $f(x)$ é conhecida como o *integrando* (a função a ser integrada), e a parte dx – semelhante a dx no operador de diferenciação d/dx – nos lembra que a operação deve ser executada em relação à variável x . Contudo, podemos também considerar $f(x) \, dx$ como uma entidade única e interpretá-la como a diferencial da função primitiva $F(x)$ [isto é, $dF(x) = f(x) \, dx$]. Então, o sinal de integral à frente pode ser visto como uma instrução para inverter o processo de diferenciação que deu origem à diferencial. Com essa nova notação, podemos escrever que

$$\frac{d}{dx} F(x) = f(x) \Rightarrow \int f(x) \, dx = F(x) + c \quad (14.3)$$

onde a presença de c , uma *constante de integração* arbitrária, serve para indicar a ascendência múltipla do integrando.

A integral $\int f(x) \, dx$ é conhecida, mais especificamente, como a *integral indefinida de $f(x)$* (em contraponto à *integral definida* a ser discutida na Seção 14.2), porque ela não tem nenhum valor numérico definido. Como ela é igual a $F(x) + c$, em geral seu valor variará com o valor de x (mesmo que c seja definida). Portanto, assim como uma derivada, uma integral indefinida é, em si, uma função da variável x .

Regras básicas de integração

Exatamente como há regras de diferenciação, também podemos desenvolver certas regras de integração. Como era de esperar, essas regras dependem extensamente das regras de derivação com as quais já estamos familiarizados. Por exemplo, pela seguinte fórmula de derivada para uma função de potência,

$$\frac{d}{dx} \left(\frac{x^{n+1}}{n+1} \right) = x^n \quad (n \neq -1)$$

vemos que a expressão $x^{n+1}/(n+1)$ é a função primitiva da função derivada x^n ; assim, escrevendo-as em lugar de $F(x)$ e $f(x)$ em (14.3), podemos enunciar o resultado como uma regra de integração.

Regra I (a regra da potência)

$$\int x^n dx = \frac{1}{n+1} x^{n+1} + c \quad (n \neq -1)$$

EXEMPLO 1 Encontre $\int x^3 dx$. Aqui, temos $n = 3$ e, portanto,

$$\int x^3 dx = \frac{1}{4} x^4 + c$$

EXEMPLO 2 Encontre $\int x dx$. Visto que $n = 1$, temos

$$\int x dx = \frac{1}{2} x^2 + c$$

EXEMPLO 3 Qual é $\int 1 dx$? Para encontrar essa integral, vamos lembrar que $x^0 = 1$; de modo que podemos considerar $n = 0$ na regra da potência e obter

$$\int 1 dx = x + c$$

[$\int 1 dx$ às vezes é escrita simplesmente como $\int dx$, já que $1 dx = dx$.]

EXEMPLO 4 Encontre $\int \sqrt{x^3} dx$. Visto que $\sqrt{x^3} = x^{3/2}$, temos $n = \frac{3}{2}$; por conseguinte,

$$\int \sqrt{x^3} dx = \frac{x^{5/2}}{\frac{5}{2}} + c = \frac{2}{5} \sqrt{x^5} + c$$

EXEMPLO 5 Encontre $\int \frac{1}{x^4} dx$, ($x \neq 0$). Visto que $1/x^4 = x^{-4}$, temos $n = -4$. Assim, a integral é

$$\int \frac{1}{x^4} dx = \frac{x^{-4+1}}{-4+1} + c = -\frac{1}{3x^3} + c$$

Note que a correção dos resultados de integração sempre pode ser verificada por diferenciação; se o processo de integração estiver correto, a derivada da integral deve ser igual ao integrando.

Já demonstramos que as fórmulas de derivadas para funções exponenciais e logarítmicas simples são

$$\frac{d}{dx} e^x = e^x \quad \text{e} \quad \frac{d}{dx} \ln x = \frac{1}{x} \quad (x > 0)$$

Dessas, surgem duas outras regras básicas de integração.

Regra II (a regra exponencial)

$$\int e^x dx = e^x + c$$

Regra III (a regra logarítmica)

$$\int \frac{1}{x} dx = \ln x + c \quad (x > 0)$$

É de interesse que o integrando envolvido na Regra III seja $1/x = x^{-1}$, que é uma forma especial da função de potência x^n com $n = -1$. Esse integrando, em particular, é inadmissível sob a regra da potência, mas agora é devidamente resolvido pela regra logarítmica.

Como exposto, a regra logarítmica está sob a restrição $x > 0$ porque não existem logaritmos para valores não-positivos de x . Uma formulação mais geral da regra, que pode resolver também os valores negativos de x , é

$$\int \frac{1}{x} dx = \ln |x| + c \quad (x \neq 0)$$

que também implica que $(d/dx) \ln |x| = 1/x$, exatamente como $(d/dx) \ln x = 1/x$. É bom que você se convença de que a substituição de x (com a restrição $x > 0$) por $|x|$ (com a restrição $x \neq 0$) não introduz absolutamente nenhum vício na fórmula.

Além disso, por questão de notação, devemos destacar que a integral $\int \frac{1}{x} dx$ às vezes também é escrita como $\int \frac{dx}{x}$.

Como variantes das Regras II e III, temos também as duas regras seguintes.

Regra IIa

$$\int f'(x) e^{f(x)} dx = e^{f(x)} + c$$

Regra IIIa

$$\int \frac{f'(x)}{f(x)} dx = \ln f(x) + c \quad [f(x) > 0]$$

ou $\ln |f(x)| + c \quad [f(x) \neq 0]$

As bases para essas duas regras podem ser encontradas nas regras de derivadas em (10.20).

Regras de operação

As três regras precedentes ilustram muito bem o espírito subjacente a todas as regras de integração. Cada regra sempre corresponde a uma certa fórmula de derivada. Além disso, uma constante arbitrária é sempre anexada ao final (embora essa constante passe de arbitrária a definida mais tarde pela utilização de uma dada condição de fronteira) para indicar que toda uma família de funções primitivas pode dar origem ao integrando dado.

Para podermos tratar com integrandos mais complicados, entretanto, veremos também que as duas regras de operação referentes a integrais apresentadas a seguir são úteis.

Regra IV (a integral de uma soma) A integral da soma de um número finito de funções é a soma das integrais dessas funções. Para o caso de duas funções, isso significa que

$$\int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx$$

Essa regra é uma consequência natural do fato de que

$$\underbrace{\frac{d}{dx} [F(x) + G(x)]}_A = \underbrace{\frac{d}{dx} F(x) + \frac{d}{dx} G(x)}_B = \underbrace{f(x) + g(x)}_C$$

Considerando que $A = C$, podemos escrever, com base em (14.3),

$$\int [f(x) + g(x)] dx = F(x) + G(x) + c \quad (14.4)$$

Mas, como $B = C$, segue-se que

$$\int f(x) dx = F(x) + c_1 \quad \text{e} \quad \int g(x) dx = G(x) + c_2$$

Assim, podemos obter (por adição)

$$\int f(x) dx + \int g(x) dx = F(x) + G(x) + c_1 + c_2 \quad (14.5)$$

Uma vez que as constantes c , c_1 e c_2 têm valor arbitrário, podemos fazer $c = c_1 + c_2$. Então, os lados direitos de (14.4) e (14.5) ficam iguais e, como consequência, seus lados esquerdos também devem ser iguais. Isso comprova a Regra IV.

EXEMPLO 6

Encontre $\int (x^3 + x + 1) dx$. Pela Regra IV, essa integral pode ser expressa como uma soma de três integrais: $\int x^3 dx + \int x dx + \int 1 dx$. Visto que os valores dessas três integrais já foram encontrados anteriormente nos Exemplos 1, 2 e 3, podemos simplesmente combinar esses resultados para obter

$$\int (x^3 + x + 1) dx = \left(\frac{x^4}{4} + c_1 \right) + \left(\frac{x^2}{2} + c_2 \right) + (x + c_3) = \frac{x^4}{4} + \frac{x^2}{2} + x + c$$

Na resposta final, agrupamos as três constantes com índices em uma única constante c .

Como prática geral, todas as constantes de integração aditivas arbitrárias que surgem durante o processo sempre podem ser combinadas em uma única constante arbitrária na resposta final.

EXEMPLO 7

Encontre $\int \left(2e^{2x} + \frac{14x}{7x^2 + 5} \right) dx$. Pela Regra IV, podemos integrar os dois termos aditivos que aparecem no integrando em separado, e então somar os resultados. Visto que o termo $2e^{2x}$ está no formato de $f'(x)e^{f(x)}$ na Regra IIa, com $f(x) = 2x$, a integral é $e^{2x} + c_1$. De modo semelhante, o outro termo, $14x/(7x^2 + 5)$, toma a forma de $f'(x)/f(x)$, com $f(x) = 7x^2 + 5 > 0$. Assim, pela Regra IIIa, a integral é $\ln(7x^2 + 5) + c_2$. Por conseguinte, podemos escrever

$$\int \left(2e^{2x} + \frac{14x}{7x^2 + 5} \right) dx = e^{2x} + \ln(7x^2 + 5) + c$$

onde combinamos c_1 e c_2 em uma única constante arbitrária c .

Regra V (a integral de um múltiplo) A integral de k vezes um integrando (sendo k uma constante) é k vezes a integral daquele integrando. Em símbolos,

$$\int kf(x) \, dx = k \int f(x) \, dx$$

Em termos operacionais, essa regra equivale a dizer que uma constante multiplicativa pode ser “fatorada para fora” do sinal de integral. (*Atenção:* Um termo *variável* não pode ser fatorado dessa mesma maneira!) Para comprovar essa regra (para o caso em que k é um inteiro), lembramos que k vezes $f(x)$ significa apenas somar $f(x)$ k vezes; por conseguinte, pela Regra IV,

$$\begin{aligned} \int kf(x) \, dx &= \int \underbrace{[f(x) + f(x) + \dots + f(x)]}_{k \text{ termos}} \, dx \\ &= \int f(x) \, dx + \int f(x) \, dx + \dots + \int f(x) \, dx = k \int f(x) \, dx \end{aligned}$$

EXEMPLO 8 Encontre $\int -f(x) \, dx$. Aqui, $k = -1$, e, assim,

$$\int -f(x) \, dx = - \int f(x) \, dx$$

Isto é, a integral da negativa de uma função é a negativa da integral daquela função.

EXEMPLO 9 Encontre $\int 2x^2 \, dx$. Fatorando o 2 para fora do sinal de integral e aplicando a Regra I, temos

$$\int 2x^2 \, dx = 2 \int x^2 \, dx = 2 \left(\frac{x^3}{3} + c_1 \right) = \frac{2}{3} x^3 + c$$

EXEMPLO 10 Encontre $\int 3x^2 \, dx$. Nesse caso, a fatoração da constante multiplicativa para fora do sinal de integral resulta em

$$\int 3x^2 \, dx = 3 \int x^2 \, dx = 3 \left(\frac{x^3}{3} + c_1 \right) = x^3 + c$$

Note que, comparando com o exemplo anterior, o termo x^3 na resposta final não é precedido de nenhuma expressão fracionária. Esse resultado elegante deve-se ao fato de que 3 (a constante multiplicativa do integrando) é exatamente igual a 2 (a potência da função) mais 1. Referindo-nos à regra da potência (Regra I), vemos que, em tal caso, a constante multiplicativa $(n + 1)$ cancelará a fração $1/(n + 1)$ resultando, por conseguinte, em $(x^{n+1} + c)$.

Em geral, sempre que tivermos uma expressão $(n + 1)x^n$ como integrando, não há realmente nenhuma necessidade de fatorar a constante $(n + 1)$ para fora do sinal de integral e depois integrar x^n ; em vez disso, podemos escrever imediatamente $x^{n+1} + c$ como resposta.

EXEMPLO 11 Encontre $\int \left(5e^x - x^{-2} + \frac{3}{x} \right) dx$, ($x \neq 0$). Este exemplo ilustra ambas as Regras IV e V; na verdade, ele ilustra também as três primeiras regras:

$$\int \left(5e^x - \frac{1}{x^2} + \frac{3}{x} \right) dx = 5 \int e^x dx - \int x^{-2} dx + 3 \int \frac{1}{x} dx \quad [\text{pelas Regras IV e V}]$$

$$\begin{aligned}
 &= (5e^x + c_1) - \left(\frac{x^{-1}}{-1} + c_2 \right) + (3 \ln |x| + c_3) \\
 &= 5e^x + \frac{1}{x} + 3 \ln |x| + c
 \end{aligned}$$

Mais uma vez, a correção do resultado pode ser verificada por diferenciação.

Regras que envolvem substituição

Agora introduziremos mais duas regras de integração que buscam simplificar o processo de integração, quando as circunstâncias são adequadas, pela substituição da variável de integração original. Sempre que a variável de integração recém-introduzida fizer com que o processo de integração fique mais fácil do que com a variável antiga, essas regras serão úteis.

Regra VI (a regra da substituição) A integral de $f(u) (du/dx)$ em relação à variável x é a integral de $f(u)$ em relação à variável u :

$$\int f(u) \frac{du}{dx} dx = \int f(u) du = F(u) + c$$

onde a operação $\int du$ foi substituída pela operação $\int dx$.

Essa regra, a contraparte da regra da cadeia no cálculo integral, pode ser comprovada por meio da própria regra da cadeia. Dada a função $F(u)$, onde $u = u(x)$, a regra da cadeia determina que

$$\frac{d}{dx} F(u) = \frac{d}{du} F(u) \frac{du}{dx} = F'(u) \frac{du}{dx} = f(u) \frac{du}{dx}$$

Visto que $f(u) (du/dx)$ é a derivada de $F(u)$, deduz-se, de (14.3), que a integral (antiderivada) da primeira deve ser

$$\int f(u) \frac{du}{dx} dx = F(u) + c$$

Você talvez tenha notado que, na verdade, esse resultado também é conseguido pelo *cancelamento* das duas expressões dx à esquerda.

EXEMPLO 12 Encontre $\int 2x(x^2 + 1) dx$. A resposta pode ser obtida primeiramente multiplicando o integrando:

$$\int 2x(x^2 + 1) dx = \int (2x^3 + 2x) dx = \frac{x^4}{2} + x^2 + c$$

mas, agora, vamos resolver pela regra de substituição. Seja $u = x^2 + 1$; então $du/dx = 2x$, ou $dx = du/2x$. Substituindo $du/2x$ por dx , temos

$$\begin{aligned}
 \int 2x(x^2 + 1) dx &= \int 2xu \frac{du}{2x} = \int u du = \frac{u^2}{2} + c_1 \\
 &= \frac{1}{2} (x^4 + 2x^2 + 1) + c_1 = \frac{1}{2} x^4 + x^2 + c
 \end{aligned}$$

onde $c = \frac{1}{2} + c_1$. Também é possível obter a mesma resposta substituindo du/dx por $2x$ (em vez de $du/2x$ por dx).

EXEMPLO 13 Encontre $\int 6x^2(x^3 + 2)^{99} dx$. Não é fácil simplificar o integrando desse exemplo por multiplicação e, assim, a regra de substituição agora tem uma oportunidade melhor de demonstrar sua eficiência. Seja $u = x^3 + 2$; então $du/dx = 3x^2$, de modo que

$$\begin{aligned}\int 6x^2(x^3 + 2)^{99} dx &= \int \left(2 \frac{du}{dx}\right) u^{99} dx = \int 2u^{99} du \\ &= \frac{2}{100} u^{100} + c = \frac{1}{50} (x^3 + 2)^{100} + c\end{aligned}$$

EXEMPLO 14 Encontre $\int 8e^{2x+3} dx$. Seja $u = 2x + 3$; então $du/dx = 2$, ou $dx = du/2$. Daí,

$$\int 8e^{2x+3} dx = \int 8 e^u \frac{du}{2} = 4 \int e^u du = 4e^u + c = 4e^{2x+3} + c$$

Como esses exemplos mostram, essa regra é útil sempre que pudermos – pela escolha judiciosa de uma função $u = u(x)$ – expressar o integrando (uma função de x) como o produto de $f(u)$ (uma função de u) e du/dx (a derivada da função u que escolhemos). Contudo, como ilustrado pelos dois últimos exemplos, essa regra também pode ser usada quando o integrando original puder ser transformado em um múltiplo constante de $f(u) (du/dx)$. Isso não afetaria a aplicabilidade porque o multiplicador constante pode ser fatorado para fora do sinal de integral, o que deixaria, então, um integrando da forma $f(u) (du/dx)$, como requer a regra de substituição. Contudo, quando a substituição de variáveis resultar em uma *variável* múltipla de $f(u) (du/dx)$, digamos, x vezes esta última, a fatoração não é permitida e essa regra não será útil. De fato, não existe nenhuma fórmula geral que dá a integral de um produto de duas funções em termos das integrais isoladas dessas funções; e também não temos uma fórmula geral que nos dê a integral de um quociente de duas funções em termos de suas integrais isoladas. E é aqui que está a razão por que a integração, no todo, é mais difícil do que a diferenciação e, por que, quando os integrandos são complicados, é mais conveniente consultar a resposta em tabelas de fórmulas de integração já preparadas, em vez de nós mesmos fazermos a integração.

Regra VII (integração por partes) A integral de v em relação a u é igual a uv menos a integral de u em relação a v :

$$\int v du = uv - \int u dv$$

A essência dessa regra é substituir a operação $\int du$ pela operação $\int dv$.

O princípio racional que fundamenta esse resultado é relativamente simples. Primeiro, a regra de produto de diferenciais nos dá

$$d(uv) = v du + u dv$$

Se integrarmos ambos os lados da equação (isto é, integrarmos cada diferencial), obtemos uma nova equação

$$\int d(uv) = \int v du + \int u dv$$

$$\text{ou } uv = \int v du + \int u dv \quad [\text{não é preciso nenhuma constante na esquerda (por quê?)}]$$

Então, subtraindo $\int u dv$ de ambos os lados, surge o resultado mencionado anteriormente.

EXEMPLO 15 Encontre $\int x(x + 1)^{1/2} dx$. Diferentemente dos exemplos 12 e 13, o presente exemplo não se presta ao tipo de substituição usado na Regra VI. (Por quê?) Contudo, podemos considerar que a integral dada está na forma de $\int v du$, e aplicar a Regra VII. Com essa finalidade, seja $v = x$, im-

plicando $dv = dx$, e também seja $u = \frac{2}{3}(x+1)^{3/2}$, de modo que $du = (x+1)^{1/2} dx$. Então, podemos constatar que a integral é

$$\begin{aligned}\int x(x+1)^{1/2} dx &= \int v du = uv - \int u dv \\ &= \frac{2}{3}(x+1)^{3/2}x - \int \frac{2}{3}(x+1)^{3/2} dx \\ &= \frac{2}{3}(x+1)^{3/2}x - \frac{4}{15}(x+1)^{5/2} + c\end{aligned}$$

EXEMPLO 16 Encontre $\int \ln x dx$, ($x > 0$). Não podemos aplicar a regra logarítmica aqui, porque essa regra trata do integrando $1/x$, e não $\ln x$. Tampouco podemos usar a Regra VI. Mas, se escolhermos $v = \ln x$, implicando $dv = (1/x) dx$, e também $u = x$, de modo que $du = dx$, então a integração pode ser realizada como se segue:

$$\begin{aligned}\int \ln x dx &= \int v du = uv - \int u dv \\ &= x \ln x - \int dx = x \ln x - x + c = x(\ln x - 1) + c\end{aligned}$$

EXEMPLO 17 Encontre $\int xe^x dx$. Nesse caso, simplesmente escolhemos $v = x$, e $u = e^x$, de modo que $dv = dx$ e $du = e^x dx$. Então, aplicando a Regra VII, temos

$$\begin{aligned}\int xe^x dx &= \int v du = uv - \int u dv \\ &= e^x x - \int e^x dx = e^x x - e^x + c = e^x(x-1) + c\end{aligned}$$

É claro que a validade desse resultado, bem como a dos exemplos precedentes, pode ser verificada imediatamente por diferenciação.

EXERCÍCIO 14.2

1. Encontre o seguinte:

(a) $\int 16x^{-3} dx \quad (x \neq 0)$

(b) $\int 9x^8 dx$

(c) $\int (x^5 - 3x) dx$

(d) $\int 2e^{-2x} dx$

(e) $\int \frac{4x}{x^2+1} dx$

(f) $\int (2ax+b)(ax^2+bx)^7 dx$

2. Encontre:

(a) $\int 13e^x dx$

(b) $\int \left(3e^x + \frac{4}{x}\right) dx \quad (x > 0)$

(c) $\int \left(5e^x + \frac{3}{x^2}\right) dx \quad (x \neq 0)$

(d) $\int 3e^{-(2x+7)} dx$

(e) $\int 4xe^{x^2+3} dx$

(f) $\int xe^{x^2+9} dx$

3. Encontre:

(a) $\int \frac{3dx}{x} \quad (x \neq 0)$

(b) $\int \frac{dx}{x-2} \quad (x \neq 2)$

(c) $\int \frac{2x}{x^2+3} dx$

(d) $\int \frac{x}{3x^2+5} dx$

4. Encontre:

(a) $\int (x+3)(x+1)^{1/2} dx$

(b) $\int x \ln x dx \quad (x > 0)$

5. Dadas n constantes k_i (com $i = 1, 2, \dots, n$) e n funções $f_i(x)$, deduza, pelas Regras IV e V que

$$\int \sum_{i=1}^n k_i f_i(x) dx = \sum_{i=1}^n k_i \int f_i(x) dx$$

14.3 Integrais definidas

Significado de integrais definidas

Todas as integrais citadas na Seção 14.2 são da variedade *indefinida*: cada uma é uma função de uma variável e, por conseguinte, não possui nenhum valor numérico definido. Agora, para uma dada integral indefinida de uma função contínua $f(x)$,

$$\int f(x) dx = F(x) + c$$

se escolhermos dois valores de x no domínio, digamos, a e b ($a < b$), e os substituirmos sucessivamente no lado direito da equação e formarmos a diferença

$$[F(b) + c] - [F(a) + c] = F(b) - F(a)$$

obtemos um valor numérico específico, livre da variável x , bem como da constante arbitrária c . Esse valor é denominado a *integral definida* de $f(x)$ de a a b . Referimo-nos a a como o *limite inferior de integração* e a b como o *limite superior de integração*.

Para indicar os limites de integração, agora modificamos o sinal de integral de modo a formar $\int_a^b$. O cálculo da integral definida é, então, simbolizado nas seguintes etapas:

$$\int_a^b f(x) dx = F(x) \Big|_a^b = F(b) - F(a) \quad (14.6)$$

onde o símbolo $\Big|_a^b$ (também representado por $\Big|_a^b$ ou $[\cdot]_a^b$) é uma instrução para substituir x sucessivamente por b e a no resultado da integração para obter $F(b)$ e $F(a)$ e então tomar a diferença entre elas, como indicado no lado direito de (14.6). Contudo, como primeira etapa, temos de achar a integral indefinida, embora possamos omitir a constante c , visto que, de qualquer modo, esta última será descartada no processo de tomar as diferenças.

EXEMPLO 1 Calcule $\int_1^5 3x^2 dx$. Visto que a integral indefinida é $x^3 + c$, o valor dessa integral definida é

$$\int_1^5 3x^2 dx = x^3 \Big|_1^5 = (5)^3 - (1)^3 = 125 - 1 = 124$$

EXEMPLO 2 Calcule $\int_a^b ke^x dx$. Aqui, os limites de integração são dados em símbolos; por consequência, o resultado da integração também está em termos desses símbolos:

$$\int_a^b ke^x dx = ke^x \Big|_a^b = k(e^b - e^a)$$

EXEMPLO 3 Calcule $\int_0^4 \left(\frac{1}{1+x} + 2x \right) dx$, ($x \neq -1$). A integral indefinida é $\ln |1+x| + x^2 + c$; assim, a resposta é

$$\begin{aligned} \int_0^4 \left(\frac{1}{1+x} + 2x \right) dx &= \left[\ln |1+x| + x^2 \right]_0^4 \\ &= (\ln 5 + 16) - (\ln 1 + 0) \\ &= \ln 5 + 16 \quad [\text{visto que } \ln 1 = 0] \end{aligned}$$

É importante entender que ambos os limites de integração a e b referem-se a valores da variável x . Se acaso usássemos a técnica de substituição de variáveis (Regras VI e VII) durante a integração e introduzíssemos uma variável u , seria preciso tomar cuidado para *não* considerar a e b como os limites de u . O Exemplo 4 ilustrará esse ponto.

EXEMPLO 4 Calcule $\int_1^2 (2x^3 - 1)^2 (6x^2) dx$. Seja $u = 2x^3 - 1$; então, $du/dx = 6x^2$, ou $du = 6x^2 dx$. Agora, observe que, quando $x = 1$, u será 1, mas, quando $x = 2$, u será 15; em outras palavras, os limites de integração em termos da variável u devem ser 1 (inferior) e 15 (superior). Portanto, ao escrever novamente a integral dada em u , não obtemos $\int_1^2 u^2 du$, mas

$$\int_1^{15} u^2 du = \frac{1}{3} u^3 \Big|_1^{15} = \frac{1}{3} (15^3 - 1^3) = 1.124 \frac{2}{3}$$

Como alternativa, podemos, em primeiro lugar, converter novamente de u para x e então usar os limites originais de 1 e 2 para obter a resposta idêntica:

$$\left[\frac{1}{3} u^3 \right]_{u=1}^{u=15} = \left[\frac{1}{3} (2x^3 - 1)^3 \right]_{x=1}^{x=2} = \frac{1}{3} (15^3 - 1^3) = 1.124 \frac{2}{3}$$

Uma integral definida como uma área sob uma curva

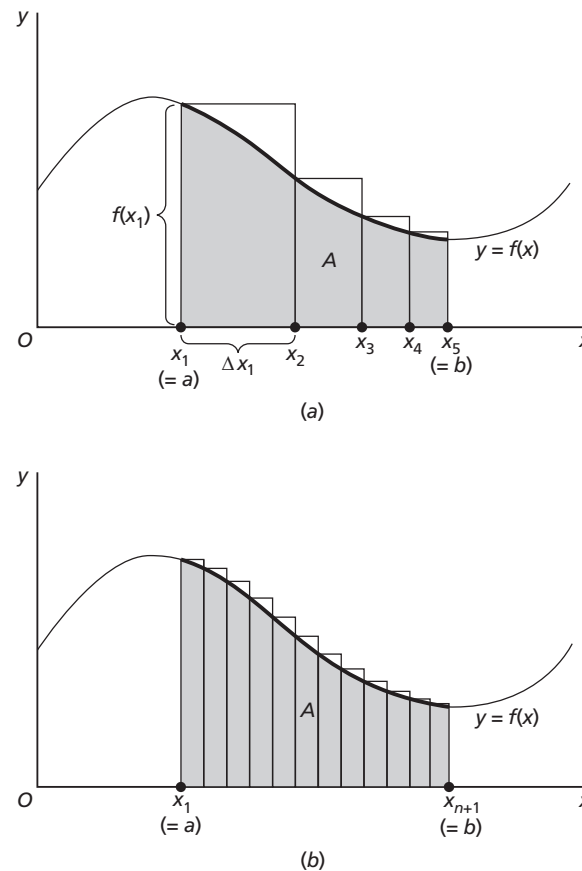
Cada integral definida tem um valor definido. Em termos geométricos, esse valor pode ser interpretado como uma área específica sob uma dada curva.

Apresentamos o gráfico de uma função contínua $y = f(x)$ na Figura 14.1. Se quisermos medir a área (sombreada) A limitada pela curva e pelo eixo x entre os dois pontos a e b no domínio, podemos proceder da seguinte maneira. Em primeiro lugar, dividimos o intervalo $[a, b]$ em n subintervalos (não necessariamente de comprimentos iguais). Quatro desses intervalos estão desenhados na Figura 14.1a – isto é, $n = 4$ –, sendo o primeiro $[x_1, x_2]$ e o último $[x_4, x_5]$. Visto que cada um desses intervalos representa uma variação em x , podemos nos referir a eles como $\Delta x_1, \dots, \Delta x_4$, respectivamente. Agora, vamos construir quatro blocos retangulares nos subintervalos de modo que a altura de cada bloco seja igual ao valor mais alto da função alcançado naquele bloco (que, aqui, por acaso, ocorre na fronteira esquerda de cada retângulo). Assim, o primeiro bloco tem altura $f(x_1)$ e largura Δx_1 e, em geral, o i -ésimo bloco tem altura $f(x_i)$ e largura Δx_i . A área total A^* desse conjunto de blocos é a soma

$$A^* = \sum_{i=1}^n f(x_i) \Delta x_i \quad (n = 4 \text{ na Figura 14.1a})$$

No entanto, é óbvio que essa área *não* é a área sob a curva que procuramos, mas apenas uma aproximação muito grosseira da área.

FIGURA 14.1



O que faz com que A^* se desvie do valor verdadeiro de A é a porção não sombreada dos blocos retangulares; elas fazem de A^* uma *superestimativa* de A . Contudo, se o tamanho da porção não sombreada puder ser reduzido de modo a se aproximar de zero, o valor aproximado de A^* se aproximará equivalentemente do verdadeiro valor de A . Esse resultado se materializará quando tentarmos uma segmentação cada vez mais fina do intervalo $[a, b]$, de modo que n aumenta e Δx_i diminui indefinidamente. Então os blocos ficarão cada vez mais delgados (se bem que mais numerosos) e a projeção que ultrapassa a curva diminuirá, como podemos ver na Figura 14.1b. Levada ao limite, essa operação de “adelgaçamento” resulta em

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x_i = \lim_{n \rightarrow \infty} A^* = \text{área } A \quad (14.7)$$

contanto que esse limite exista. (E existe, no presente caso.) Na verdade, essa equação constitui a definição formal de uma área sob uma curva.

A expressão de somatório em (14.7), $\sum_{i=1}^n f(x_i) \Delta x_i$, guarda uma certa semelhança com a expressão da integral definida $\int_a^b f(x) dx$. De fato, a última é baseada na primeira. A substituição de Δx_i pela diferencial dx é feita segundo o mesmo critério adotado quando discutimos “aproximação” na Seção 8.1. Assim, reescrevemos $f(x_i) \Delta x_i$ como $f(x) dx$. E o sinal de somatório? A notação $\sum_{i=1}^n$ representa a soma de um número *finito* de termos. Quando fazemos $n \rightarrow \infty$ e tomamos o limite daquela soma, a notação comum para tal operação é bastante desajeitada. Assim, precisamos de uma substituta mais simples. Essa substituta é $\int_a^b$, onde o símbolo alongado S também indica

uma soma, e onde a e b (exatamente como $i = 1$ e n) servem para especificar os limites inferior e superior dessa soma. Em resumo, a integral definida é uma forma abreviada para a expressão de limite de uma soma em (14.7). Isto é,

$$\int_a^b f(x) \, dx \equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x_i = \text{área } A$$

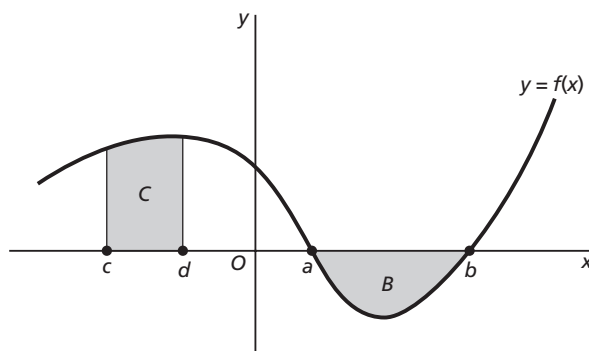
Assim, a integral definida citada (denominada uma *integral de Riemann*) agora tem uma conotação de *área*, bem como de *soma*, porque $\int_a^b$ é a contraparte contínua do conceito discreto de $\sum_{i=1}^n$.

Na Figura 14.1, tentamos aproximar a área A reduzindo sistematicamente uma *superestimativa* de A^* para segmentação mais fina do intervalo $[a, b]$. O limite resultante da soma das áreas dos blocos é denominado a *integral superior* – uma aproximação por cima. Também poderíamos ter aproximado a área A por baixo, formando blocos retangulares inscritos pela curva em vez de se projetarem para além dela (veja Exercício 14.3-3). A área total A^{**} desse novo conjunto de blocos será uma *subestimativa* de A , mas, à medida que a segmentação de $[a, b]$ se tornar cada vez mais fina, novamente acharemos $\lim_{n \rightarrow \infty} A^{**} = A$. Esse limite da soma de áreas de blocos que acabamos de citar é denominado a *integral inferior*. Se, e somente se, a integral superior e a integral inferior tiverem valores iguais, então a integral de $\int_a^b f(x) \, dx$ é definida e diz-se que a função $f(x)$ é *integrável segundo Riemann*. Existem teoremas que especificam as condições sob as quais uma função $f(x)$ é integrável. Segundo o teorema fundamental do cálculo, uma função é integrável em $[a, b]$ se for contínua nesse intervalo. Portanto, enquanto trabalharmos com funções contínuas, não deveremos ter nenhuma preocupação quanto a isso.

Um outro ponto pode ser notado. Conquanto a área A na Figura 14.1 esteja inteiramente sob uma porção decrescente da curva $y = f(x)$, a igualdade conceitual entre uma integral definida e uma área também é válida para porções da curva cuja inclinação é ascendente. De fato, ambos os tipos de inclinação podem se apresentar simultaneamente; por exemplo, podemos calcular $\int_0^b f(x) \, dx$ como a área sob a curva na Figura 14.1 que está acima da linha Ob .

Note que, se calcularmos a área B na Figura 14.2 pela integral definida $\int_a^b f(x) \, dx$, a resposta dará negativa porque a altura de cada bloco retangular envolvido nessa área é negativa. Isso dá origem à idéia de uma *área negativa*, uma área que fica *abaixo* do eixo x e *acima* de uma curva dada. Portanto, caso estejamos interessados no valor numérico, e não no valor algébrico de tal área, devemos tomar o valor absoluto da integral definida relevante. A área $C = \int_c^d f(x) \, dx$, por outro lado, tem sinal positivo ainda que esteja na região negativa do eixo x ; isso porque cada bloco retangular tem uma altura positiva, bem como uma largura positiva quando passamos de c para d . De tudo isso, fica clara a implicação que a permuta dos dois limites de integração, por inverter a direção do movimento, alteraria o sinal de Δx_i e o da integral definida. Aplicando isso à área B , vemos que a integral definida $\int_b^a f(x) \, dx$ (de b para a) dará a negativa da área B ; isso medirá o valor numérico dessa área.

FIGURA 14.2



Algumas propriedades de integrais definidas

A discussão no parágrafo precedente nos leva à seguinte propriedade de integrais definidas.

Propriedade I A permuta de limites de integração muda o sinal da integral definida:

$$\int_b^a f(x) \, dx = - \int_a^b f(x) \, dx$$

Isso pode ser comprovado como se segue:

$$\int_b^a f(x) \, dx = F(a) - F(b) = - [F(b) - F(a)] = - \int_a^b f(x) \, dx$$

Integrais definidas também possuem algumas outras propriedades interessantes.

Propriedade II Uma integral definida tem valor zero quando os dois limites de integração forem idênticos:

$$\int_a^a f(x) \, dx = F(a) - F(a) = 0$$

Segundo a interpretação de “área”, isso significa que a área (sob uma curva) acima de qualquer *ponto* único no domínio é nula. E é assim que deve ser, porque, em cima de um ponto no eixo x , podemos desenhar apenas uma *reta* (unidimensional), nunca uma *área* (bidimensional).

Propriedade III Uma integral definida pode ser expressa como uma soma de um número finito de subintegrais, como se segue:

$$\int_a^d f(x) \, dx = \int_a^b f(x) \, dx + \int_b^c f(x) \, dx + \int_c^d f(x) \, dx \quad (a < b < c < d)$$

Somente três subintegrais são mostradas nessa equação, mas a extensão para o caso de n subintegrais também é válida. Essa propriedade às vezes é descrita como *aditividade*.

Em termos de área, isso significa que a área (sob a curva) que está acima do intervalo $[a, d]$ no eixo x pode ser obtida somando as áreas que estão acima dos subintervalos no conjunto $\{[a, b], [b, c], [c, d]\}$. Note que, visto que estamos lidando com intervalos fechados, cada um dos pontos de fronteira b e c foi incluído em *duas* áreas. Isso não é contagem dupla? De fato, é. Mas, felizmente, nenhum dano é causado, porque, pela Propriedade II, a área acima de um único ponto é zero, de modo que a contagem dupla não produz nenhum efeito sobre o cálculo. Porém, não é preciso nem dizer que a contagem dupla de qualquer *intervalo* nunca é permitida.

Mencionamos anteriormente que todas as funções contínuas são integráveis segundo Riemann. Agora, pela Propriedade III, podemos também achar as integrais definidas (áreas) de certas funções descontínuas. Considere a função degrau na Figura 14.3a. A despeito da descontinuidade no ponto b no intervalo $[a, c]$, podemos encontrar a área sombreada a partir da soma

$$\int_a^b f(x) \, dx + \int_b^c f(x) \, dx$$

O mesmo se aplica à curva na Figura 14.3b.

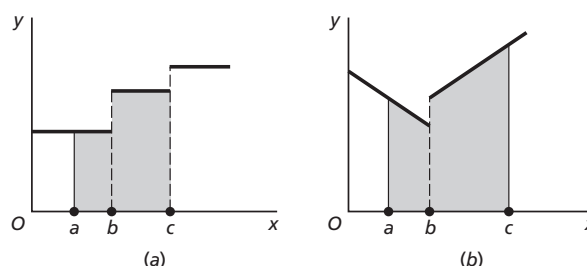
Propriedade IV

$$\int_a^b -f(x) \, dx = - \int_a^b f(x) \, dx$$

Propriedade V

$$\int_a^b kf(x) \, dx = k \int_a^b f(x) \, dx$$

FIGURA 14.3

**Propriedade VI**

$$\int_a^b [f(x) + g(x)] dx = \int_a^b f(x) dx + \int_a^b g(x) dx$$

Propriedade VII (integração por partes) Dadas $u(x)$ e $v(x)$,

$$\int_{x=a}^{x=b} v du = uv \Big|_{x=a}^{x=b} - \int_{x=a}^{x=b} u dv$$

Essas últimas quatro propriedades, todas emprestadas das regras da integração indefinida, devem prescindir de maiores explicações.

Um outro modo de ver a integral indefinida

Apresentamos a integral definida via a anexação de dois limites de integração a uma integral indefinida. Agora que sabemos o significado da integral definida, vamos ver como podemos reverter desta última para a integral indefinida.

Suponha que, em vez de fixar o limite superior da integração em b , permitamos que ele seja uma variável, designado simplesmente por x . Então, a integral assumirá a forma

$$\int_a^x f(x) dx = F(x) - F(a)$$

a qual, sendo agora uma função de x , denota uma área *variável* sob a curva de $f(x)$. Mas, visto que o último termo à direita é uma constante, essa integral deve ser um membro da família das funções primitivas de $f(x)$, que denotamos anteriormente por $F(x) + c$. Se estabelecermos $c = -F(a)$, então a integral acima torna-se exatamente a integral indefinida $\int f(x) dx$.

Por conseguinte, desse ponto de vista, podemos considerar que o símbolo $\int$ significa o mesmo que $\int_a^x$, contanto que fique entendido que, na última versão do símbolo, o limite inferior de integração está relacionado à constante de integração pela equação $c = -F(a)$.

EXERCÍCIO 14.3

1. Calcule as seguintes:

(a) $\int_1^3 \frac{1}{2} x^2 dx$

(d) $\int_2^4 (x^3 - 6x^2) dx$

(b) $\int_0^1 x(x^2 + 6) dx$

(e) $\int_{-1}^1 (ax^2 + bx + c) dx$

(c) $\int_1^3 3\sqrt{x} dx$

(f) $\int_4^2 x^2 \left(\frac{1}{3} x^3 + 1 \right) dx$

2. Calcule as seguintes:

(a) $\int_1^2 e^{-2x} dx$

(c) $\int_2^3 (e^{2x} + e^x) dx$

(b) $\int_{-1}^{e-2} \frac{dx}{x+2}$

(d) $\int_e^6 \left(\frac{1}{x} + \frac{1}{1+x} \right) dx$

 3. Na Figura 14.1a, tome o valor mais baixo da função obtido em cada subintervalo como a altura do bloco retangular, isto é, tome $f(x_2)$ em vez de $f(x_1)$ como a altura do primeiro bloco, embora ainda conservando Δx_1 como sua largura, e faça o mesmo para os outros blocos.

 (a) Escreva uma expressão de somatório para a área total A^{**} dos novos retângulos.

 (b) A área A^{**} superestima ou subestima a área desejada A ?

 (c) A área A^{**} tenderia a se aproximar ou a se afastar mais ainda de A se partirmos para uma segmentação mais fina de $[a, b]$? (*Sugestão:* Experimente um diagrama).

 (d) No limite, quando o número n de subintervalos se aproxima de ∞ , o valor aproximado de A^{**} se aproximaria do valor verdadeiro de A , exatamente como o valor aproximado de A^* o fez?

 (e) O que você pode concluir de (a) a (d) sobre a integrabilidade segundo Riemann da função $f(x)$ na Figura 14.1a?

 4. Diz-se que a integral definida $\int_a^b f(x) dx$ representa uma área sob uma curva. Essa curva se refere ao gráfico do integrando $f(x)$, ou da função primitiva $F(x)$? Se desenharmos o gráfico da função $F(x)$, como podemos mostrar sobre ele a integral definida dada – por uma área, por um segmento de reta, ou por um ponto?

 5. Verifique se a constante c pode ser expressa de modo equivalente como uma integral definida:

(a) $c \equiv \int_0^{\frac{bc}{d}} \frac{dx}{d}$

(b) $c \equiv \int_0^c 1 dt$

14.4 Integrais impróprias

Diz-se que certas integrais são “impróprias”. Discutiremos resumidamente duas de suas variedades.

Limites de integração infinitos

Quando temos integrais definidas da forma

$$\int_a^\infty f(x) dx \quad \text{e} \quad \int_{-\infty}^b f(x) dx$$

com um limite de integração infinito, nos referimos a elas como *integrais impróprias*. Nesses casos, não é possível avaliar as integrais como, respectivamente,

$$F(\infty) - F(a) \quad \text{e} \quad F(b) - F(-\infty)$$

porque ∞ não é um número e, portanto, não pode substituir x na função $F(x)$. Em vez disso, devemos recorrer mais uma vez ao conceito de limites.

A primeira integral imprópria que citamos pode ser definida como o limite de uma outra integral (própria) quando o limite superior de integração desta última tender a ∞ ; isto é,

$$\int_a^\infty f(x) dx \equiv \lim_{b \rightarrow \infty} \int_a^b f(x) dx \quad (14.8)$$

Se esse limite existir, diz-se que a integral imprópria é convergente (ou converge), e o processo de cálculo do limite dará o valor da integral. Se o limite não existir, diz-se que a integral imprópria é divergente e, na verdade, não tem significado. Pelo mesmo raciocínio, podemos definir

$$\int_{-\infty}^b f(x) dx \equiv \lim_{a \rightarrow -\infty} \int_a^b f(x) dx \quad (14.8')$$

com o mesmo critério de convergência e divergência.

EXEMPLO 1 Calcule $\int_1^{\infty} \frac{dx}{x^2}$. Primeiro, notamos que

$$\int_1^b \frac{dx}{x^2} = \left[-\frac{1}{x} \right]_1^b = -\frac{1}{b} + 1$$

Portanto, segundo (14.8), a integral desejada é

$$\int_1^{\infty} \frac{dx}{x^2} = \lim_{b \rightarrow \infty} \int_1^b \frac{dx}{x^2} = \lim_{b \rightarrow \infty} \left(-\frac{1}{b} + 1 \right) = 1$$

Essa integral imprópria converge e seu valor é 1.

Uma vez que a expressão de limite é trabalhosa de escrever, há quem prefira omitir a notação “lim” e escrever simplesmente

$$\int_1^{\infty} \frac{dx}{x^2} = \left[-\frac{1}{x} \right]_1^{\infty} = 0 + 1 = 1$$

Entretanto, mesmo quando escrita dessa forma, a integral imprópria deve ser interpretada tendo em mente o conceito de limite.

Em termos gráficos, essa integral imprópria ainda tem a conotação de uma área. Mas, visto que é permitido que o limite superior de integração assuma valores cada vez maiores nesse caso, a fronteira do lado direito deve ser estendida indefinidamente na direção leste, como mostra a Figura 14.4a. A despeito disso, podemos considerar que a área tem o valor definido (limite) de 1.

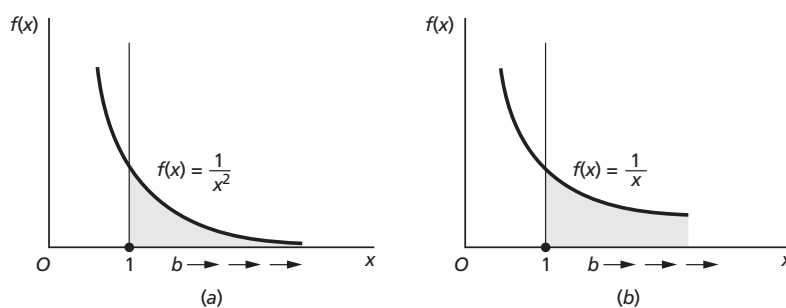
EXEMPLO 2 Calcule $\int_1^{\infty} \frac{dx}{x}$. Como antes, em primeiro lugar encontramos

$$\int_1^b \frac{dx}{x} = \ln x \Big|_1^b = \ln b - \ln 1 = \ln b$$

Quando fazemos $b \rightarrow \infty$, temos, por (10.16'), $\ln b \rightarrow \infty$. Assim, a integral imprópria dada é divergente.

A Figura 14.4b mostra o gráfico da função $1/x$, bem como a área correspondente à integral dada. A extensão indefinida da fronteira do lado direito na direção leste resultará, desta vez, em uma área infinita, mesmo que a forma do gráfico apresente uma similaridade superficial com o gráfico da Figura 14.4a.

FIGURA 14.4



E se ambos os limites de integração forem infinitos? Uma extensão direta de (14.8) e (14.8') sugeriria a definição

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{\substack{b \rightarrow +\infty \\ a \rightarrow -\infty}} \int_a^b f(x) dx \quad (14.8'')$$

Novamente, diz-se que essa integral imprópria converge se, e somente se, o limite em questão existir.

Integrando infinito

Mesmo com limites de integração finitos, uma integral ainda pode ser imprópria se o integrando tornar-se infinito em algum ponto do intervalo de integração $[a, b]$. Para calcular tal integral, mais uma vez devemos recorrer ao conceito de um limite.

EXEMPLO 3

Calcule $\int_0^1 \frac{1}{x} dx$. Essa integral é imprópria porque, como mostra a Figura 14.4b, o integrando é infinito no limite inferior de integração ($1/x \rightarrow \infty$ as $x \rightarrow 0^+$). Por conseguinte, em primeiro lugar, devemos achar a integral

$$\int_a^1 \frac{1}{x} dx = \ln x \Big|_a^1 = \ln 1 - \ln a = -\ln a \quad [\text{para } a > 0]$$

e então calcular seu limite quando $a \rightarrow 0^+$:

$$\int_0^1 \frac{1}{x} dx \equiv \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{x} dx = \lim_{a \rightarrow 0^+} (-\ln a)$$

Uma vez que esse limite não existe (quando $a \rightarrow 0^+$, $\ln a \rightarrow -\infty$), a integral dada é divergente.

EXEMPLO 4

Calcule $\int_0^9 x^{-1/2} dx$. Quando $x \rightarrow 0^+$, o integrando $1/\sqrt{x}$ torna-se infinito; a integral é imprópria. Novamente, temos, em primeiro lugar

$$\int_a^9 x^{-1/2} dx = 2x^{1/2} \Big|_a^9 = 6 - 2\sqrt{a}$$

O limite dessa expressão quando $a \rightarrow 0^+$ é $6 - 0 = 6$. Assim, a integral dada é convergente (para 6).

A situação em que o integrando torna-se infinito no limite *superior* de integração é inteiramente análoga. Contudo, uma proposição totalmente diferente é quando ocorre um valor infinito do integrando no intervalo aberto (a, b) , e não em a ou b . Quando isso acontece, é preciso tirar proveito da aditividade de integrais definidas e primeiramente decompor a integral dada em subintegrais. Suponha que $f(x) \rightarrow \infty$ quando $x \rightarrow p$, onde p é um ponto no intervalo (a, b) ; então, pela propriedade da aditividade, temos

$$\int_a^b f(x) dx = \int_a^p f(x) dx + \int_p^b f(x) dx$$

A integral dada à esquerda pode ser considerada como convergente se, e somente se, cada subintegral tiver um limite.

EXEMPLO 5

Calcule $\int_{-1}^1 \frac{1}{x^3} dx$. o integrando tende ao infinito quando x se aproxima de zero; assim, devemos escrever a integral como a soma

$$\int_{-1}^1 x^{-3} dx = \int_{-1}^0 x^{-3} dx + \int_0^1 x^{-3} dx \quad (\text{digamos, } \equiv I_1 + I_2)$$

A integral I_1 é divergente, porque

$$\lim_{b \rightarrow 0^-} \int_{-1}^b x^{-3} dx = \lim_{b \rightarrow 0^-} \left[\frac{-1}{2} x^{-2} \right]_{-1}^b = \lim_{b \rightarrow 0^-} \left(-\frac{1}{2b^2} + \frac{1}{2} \right) = -\infty$$

Assim, podemos concluir imediatamente, sem ter de avaliar I_2 , que a integral dada é divergente.

EXERCÍCIO 14.4

- Verifique as integrais definidas dadas nos Exercícios 14.3-1 e 14.3-2 para determinar se alguma delas é imprópria. Em caso positivo, indique a qual variedade de integral imprópria cada uma pertence.
- Quais das seguintes integrais são impróprias e por quê?

(a) $\int_0^{\infty} e^{-rt} dt$	(d) $\int_{-\infty}^0 e^{rt} dt$
(b) $\int_2^3 x^4 dx$	(e) $\int_1^5 \frac{dx}{x-2}$
(c) $\int_0^1 x^{-2/3} dx$	(f) $\int_{-3}^4 6 dx$
- Calcule todas as integrais *impróprias* no Problema 2.
- Calcule a integral I_2 do Exemplo 5 e mostre que ela também é divergente.
- (a) Construa o gráfico da função $y = ce^{-t}$ para t não-negativo, ($c > 0$), e sombreie a área sob a curva.
 (b) Escreva uma expressão matemática para essa área e determine se é uma área finita.

14.5 Algumas aplicações econômicas de integrais

Integrais são usadas em análise econômica de muitas maneiras. Vamos ilustrar algumas aplicações simples nesta seção e mostrar a aplicação do modelo de crescimento de Domar na Seção 14.6.

De uma função marginal para uma função total

Dada uma função total (por exemplo, uma função custo total), o processo de diferenciação pode dar como resultado a função marginal (por exemplo, a função custo marginal). Como o processo de integração é o oposto do processo de diferenciação, ele deve nos capacitar, pelo caminho inverso, a inferir a função total a partir de uma função marginal dada.

EXEMPLO 1

Se o custo marginal (MC) de uma empresa for a seguinte função da produção, $C'(Q) = 2e^{0,2Q}$, e se o custo fixo for $C_F = 90$, encontre a função custo total $C(Q)$. Integrando $C'(Q)$ em relação a Q , constatamos que

$$\int 2e^{0,2Q} dQ = 2 \frac{1}{0,2} e^{0,2Q} + c = 10e^{0,2Q} + c \quad (14.9)$$

Esse resultado pode ser considerado como a função desejada $C(Q)$ exceto que, em vista da constante arbitrária c , a resposta aparece indeterminada. Felizmente, a informação de que $C_F = 90$ pode ser usada como uma condição inicial para definir a constante. Quando $Q = 0$, o custo total C se constituirá exclusivamente de C_F . Portanto, estabelecendo $Q = 0$ no resultado de (14.9), devemos obter um valor de 90; isto é, $10e^0 + c = 90$. Mas isso implicaria que $c = 90 - 10 = 80$. Por conseguinte, a função custo total é

$$C(Q) = 10e^{0,2Q} + 80$$

Note que, diferentemente do caso de (14.2), onde a constante arbitrária c tem o mesmo valor que o valor inicial da variável $H(0)$, no presente exemplo temos $c = 80$, mas $C(0) \equiv C_F = 90$, de modo que as duas assumem valores diferentes. Em geral, não se deve admitir de antemão que a constante arbitrária c será sempre igual ao valor inicial da função total.

EXEMPLO 2

Se a propensão marginal a poupar (MPS) for a seguinte função de renda, $S'(Y) = 0,3 - 0,1Y^{-1/2}$, e se as poupanças agregadas S forem nulas quando a renda Y for 81, encontre a função poupança $S(Y)$. Como a MPS é a derivada da função S , o problema agora pede a integração de $S'(Y)$:

$$S(Y) = \int (0,3 - 0,1Y^{-1/2}) dY = 0,3Y - 0,2Y^{1/2} + c$$

O valor específico da constante c pode ser encontrado pelo fato de que $S = 0$ quando $Y = 81$. Mesmo que, em termos estritos, esta não seja uma condição *inicial* (já que não relaciona a $Y = 0$), ainda assim a substituição dessa informação na integral precedente servirá para definir c . Visto que

$$0 = 0,3(81) - 0,2(9) + c \quad \Rightarrow \quad c = -22,5$$

a função poupança desejada é

$$S(Y) = 0,3Y - 0,2Y^{1/2} - 22,5$$

A técnica ilustrada nos Exemplos 1 e 2 pode ser estendida diretamente a outros problemas que envolvam a procura de funções totais (tais como receita total, consumo total) a partir de funções marginais dadas. Também podemos reiterar que, em problemas desse tipo, a validade da resposta (uma integral) sempre pode ser verificada por diferenciação.

Investimento e formação de capital

Formação de capital é o processo de acrescentar a um dado estoque de capital. Considerando esse processo como contínuo ao longo do tempo, podemos expressar o estoque de capital como uma função do tempo, $K(t)$, e usar a derivada dK/dt para denotar a taxa de formação de capital.[†] Mas a taxa de formação de capital no instante t é idêntica à taxa de fluxo de *investimento líquido* no instante t , denotada por $I(t)$. Assim, o estoque de capital K e o investimento líquido I estão relacionados pelas duas equações seguintes:

$$\frac{dK}{dt} \equiv I(t)$$

e

$$K(t) = \int I(t) dt = \int \frac{dK}{dt} dt = \int dK$$

[†] Por questão de notação, a derivada de uma variável em relação ao *tempo* muitas vezes também é denotada por um ponto colocado sobre a variável, tal como $\dot{K} \equiv dK/dt$. Em análise dinâmica, onde é abundante a ocorrência de derivadas em relação ao *tempo*, esse símbolo mais conciso pode dar uma contribuição substancial à simplicidade da notação. Todavia, dado que um ponto é uma marca pequenina, é fácil perdê-lo de vista ou colocá-lo no lugar errado; assim, é preciso muito cuidado ao usar esse símbolo.

A primeira das duas equações precedentes é uma identidade; ela mostra a sinonímia entre o investimento líquido e o incremento de capital. Visto que $I(t)$ é a derivada de $K(t)$, é óbvio que $K(t)$ é a integral ou antiderivada de $I(t)$, como mostra a segunda equação. A transformação do integrando na última equação também é fácil de compreender: a mudança de I para dK/dt acontece por definição e a próxima transformação acontece por cancelamento de duas diferenciais idênticas, isto é, pela regra de substituição.

Às vezes, o conceito de *investimento bruto* é usado juntamente com o de investimento líquido em um modelo. Denotando investimento bruto por I_g e investimento líquido por I , podemos relacionar um com o outro pela equação

$$I_g = I + \delta K$$

onde δ representa a taxa de depreciação do capital e δK , a taxa de *investimento de reposição*.

EXEMPLO 3 Suponha que o fluxo de investimento líquido é descrito pela equação $I(t) = 3t^{1/2}$ e que o estoque de capital inicial, no instante $t = 0$, é $K(0)$. Qual é a trajetória temporal do capital K ? Integrando $I(t)$ em relação a t , obtemos

$$K(t) = \int I(t) dt = \int 3t^{1/2} dt = 2t^{3/2} + c$$

Em seguida, fazendo $t = 0$ nas expressões da extrema esquerda e da extrema direita, achamos $K(0) = c$. Por conseguinte, a trajetória temporal de K é

$$K(t) = 2t^{3/2} + K(0) \quad (14.10)$$

Observe a semelhança básica entre os resultados em (14.10) e em (14.2'').

O conceito de integral definida entra em cena quando desejamos achar o montante de formação de capital durante um certo intervalo de tempo (e não na trajetória temporal de K). Visto que $\int I(t) dt = K(t)$, podemos escrever a integral definida

$$\int_a^b I(t) dt = K(t) \Big|_a^b = K(b) - K(a)$$

para indicar a acumulação total de capital durante o intervalo de tempo $[a, b]$. É claro que isso também representa uma área sob a curva $I(t)$. Contudo, deve-se observar que, no gráfico da função $K(t)$, essa integral definida apareceria, por sua vez, como uma distância vertical – mais especificamente, como a diferença entre as duas distâncias verticais $K(b)$ e $K(a)$. (cf. Exercício 14.3-4.)

Para avaliar essa distinção entre $K(t)$ e $I(t)$ com maior clareza, vamos enfatizar que o capital K é um conceito de *estoque*, ao passo que o investimento I é um conceito de *fluxo*. De acordo com isso, enquanto $K(t)$ nos diz qual é o *montante* de K existente em cada instante, $I(t)$ nos dá informações sobre a *taxa* de investimento (líquido) por ano (ou por período de tempo) que prevalece em cada instante. Assim, para calcular o *montante* de investimento líquido comprometido (acumulação de capital), devemos primeiramente especificar o comprimento do intervalo envolvido. Esse fato também pode ser visto quando reescrevemos a identidade $dK/dt \equiv I(t)$ como $dK \equiv I(t) dt$, que afirma que dK , o incremento em K , é baseado não somente em $I(t)$, a taxa de fluxo, mas também em dt , o tempo transcorrido. É essa necessidade de especificar o intervalo de tempo na expressão $I(t) dt$ que põe em cena a integral definida e dá origem à representação da *área* sob a curva $I(t)$ – em vez de sob a curva $K(t)$.

EXEMPLO 4 Se o investimento líquido for um fluxo constante em $I(t) = 1.000$ (dólares por ano), qual será o investimento líquido total (formação de capital) durante um ano, de $t = 0$ a $t = 1$? Obviamente, a resposta é \$1.000; esse resultado pode ser obtido formalmente da seguinte maneira:

$$\int_0^1 I(t) dt = \int_0^1 1.000 dt = 1.000t \Big|_0^1 = 1.000$$

Você pode verificar que surgirá a mesma resposta se o ano envolvido for, por outro lado, de $t = 1$ a $t = 2$.

EXEMPLO 5

Se $I(t) = 3t^{1/2}$ (milhares de dólares por ano) – um fluxo não-constante –, qual será a formação de capital durante o intervalo de tempo $[1, 4]$, isto é, durante o segundo, terceiro e quarto anos? A resposta está na integral definida

$$\int_1^4 3t^{1/2} dt = 2t^{3/2} \Big|_1^4 = 16 - 2 = 14$$

Com base nos exemplos precedentes, podemos expressar o montante de acumulação de capital durante o intervalo de tempo $[0, t]$, para qualquer taxa de investimento $I(t)$, pela integral definida

$$\int_0^t I(t) dt = K(t) - K(0)$$

A Figura 14.5 ilustra o caso do intervalo de tempo $[0, t_0]$. Vista de um modo diferente, a equação precedente resulta na seguinte expressão para a trajetória temporal $K(t)$:

$$K(t) = K(0) + \int_0^t I(t) dt$$

O montante de K em qualquer instante t é o capital inicial mais a acumulação total de capital que ocorreu desde então.

Valor presente de um fluxo de caixa

Nossa discussão anterior sobre desconto e valor presente, limitada ao caso de um *único* valor futuro V , nos levou às fórmulas de desconto

$$A = V(1 + i)^{-t} \quad [\text{caso discreto}]$$

e

$$A = Ve^{-rt} \quad [\text{caso contínuo}]$$

Agora suponha que temos uma corrente ou fluxo de valores futuros – uma série de receitas recebíveis em vários instantes ou de desembolsos de custo pagáveis em vários instantes. Como calculamos o valor presente de toda a corrente de caixa, ou fluxo de caixa?

No caso *discreto*, se admitirmos três números de receita futura R_i ($i = 1, 2, 3$) disponíveis ao final do i -ésimo ano e admitirmos uma taxa de juro de i por ano, os valores presentes de R_i serão, respectivamente,

$$R_1(1 + i)^{-1} \quad R_2(1 + i)^{-2} \quad R_3(1 + i)^{-3}$$

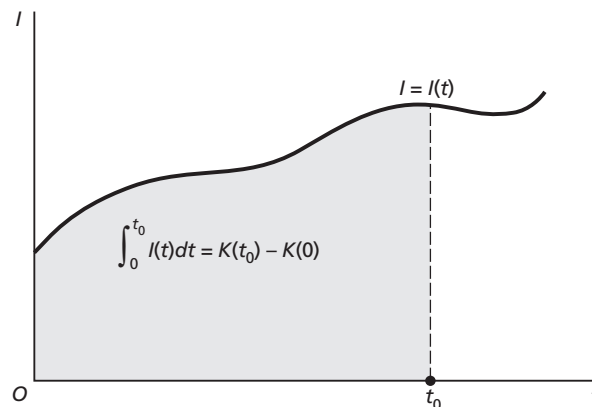


FIGURA 14.5

Segue-se que o valor presente total é a soma

$$\Pi = \sum_{t=1}^3 R_t (1+i)^{-t} \quad (14.11)$$

(Π é a letra grega maiúscula pi que, aqui, significa *presente*.) Essa fórmula difere da fórmula do valor único somente pela substituição de V por R_t e pela inserção do sinal de somatório Σ .

A idéia da soma pode ser transferida de imediato para o caso de um fluxo de caixa contínuo, mas, nesse último contexto, o símbolo Σ deve dar lugar, é claro, ao sinal de integral definida. Considere uma corrente contínua de receita a uma taxa de $R(t)$ dólares por ano. Isso significa que, em $t = t_1$, a taxa de fluxo é $R(t_1)$ dólares por ano, mas, em um outro instante $t = t_2$, a taxa será $R(t_2)$ dólares por ano – considerando t como uma variável contínua. Em qualquer instante, o montante de receita durante o intervalo $[t, t + dt]$ pode ser escrito como $R(t) dt$ [cf. a discussão anterior de $dK \equiv I(t) dt$]. Quando descontada continuamente à taxa de r por ano, seu valor presente deve ser $R(t)e^{-rt} dt$. Se, por outro lado, nosso problema for encontrar o valor presente total de um fluxo de 3 anos, nossa resposta será encontrada na seguinte integral definida:

$$\Pi = \int_0^3 R(t)e^{-rt} dt \quad (14.11')$$

Essa expressão, a versão contínua da soma em (14.11), difere da fórmula do valor único somente na substituição de V por $R(t)$ e no acréscimo do sinal da integral definida.[†]

EXEMPLO 6

Qual é o valor presente de um fluxo contínuo de receita que dura y anos à taxa constante de D dólares por ano e descontado à taxa de r por ano? Segundo (14.11'), temos

$$\begin{aligned} \Pi &= \int_0^y D e^{-rt} dt = D \int_0^y e^{-rt} dt = D \left[\frac{-1}{r} e^{-rt} \right]_0^y \\ &= \frac{-D}{r} e^{-rt} \Big|_{t=0}^{t=y} = \frac{-D}{r} (e^{-ry} - 1) = \frac{D}{r} (1 - e^{-ry}) \end{aligned} \quad (14.12)$$

Assim, Π depende de D , r e y . Se $D = \$3.000$, $r = 0,06$, e $y = 2$, por exemplo, temos

$$\Pi = \frac{3.000}{0,06} (1 - e^{-0,12}) = 50.000(1 - 0,8869) = \$5.655 \quad [\text{aproximadamente}]$$

O valor de Π naturalmente é sempre positivo, o que se deduz da positividade de D e r , bem como de $(1 - e^{-ry})$. (O número e elevado a qualquer potência negativa sempre dará um valor fracionário positivo, como podemos observar no segundo quadrante da Figura 10.3a.)

EXEMPLO 7

No problema da armazenagem de vinho na Seção 10.6, admitimos que o custo de armazenagem seria zero. Essa premissa simplificadora era necessária porque não conhecíamos um modo de calcular o valor presente de um fluxo de custo. Como isso já ficou para trás, agora estamos prontos para permitir que o negociante de vinhos incorra em custos de armazenagem.

Seja o custo de compra de uma caixa de vinhos uma quantia C , incorrida no tempo presente. Seu valor de venda (futuro), que varia com o tempo, pode ser denotado, de forma geral, como $V(t)$ – sendo seu valor presente $V(t)e^{-rt}$. Ao passo que o valor de venda representa um único valor futuro (só poderá haver uma transação de venda com essa caixa de vinhos), o custo de armazenagem é um fluxo. Admitindo que esse custo seja um fluxo constante à taxa de s dólares por ano, o valor presente total do custo de armazenagem incorrido em um total de t anos equivalerá a

[†] Podemos observar que, conquanto o índice superior do somatório e o limite superior de integração sejam idênticos em 3, o índice inferior do somatório, 1, é diferente do limite inferior de integração, 0. Isso porque, por hipótese, a primeira receita no fluxo discreto não estará disponível até $t = 1$ (final do primeiro ano), mas admite-se que o fluxo de receita no caso contínuo começa imediatamente após $t = 0$.

$$\int_0^t s e^{-rt} dt = \frac{s}{r} (1 - e^{-rt}) \quad (\text{conforme (14.12)})$$

Assim, o valor presente *líquido* – que o comerciante procuraria maximizar – pode ser expresso como

$$N(t) = V(t)e^{-rt} - \frac{s}{r} (1 - e^{-rt}) - C = \left[V(t) + \frac{s}{r} \right] e^{-rt} - \frac{s}{r} - C$$

que é uma função objetivo de uma única variável de escolha t .

Para maximizar $N(t)$, o valor de t deve ser escolhido de modo que $N'(t) = 0$. A derivada primeira é

$$\begin{aligned} N'(t) &= V'(t)e^{-rt} - r \left[V(t) + \frac{s}{r} \right] e^{-rt} \quad [\text{regra do produto}] \\ &= [V'(t) - rV(t) - s]e^{-rt} \end{aligned}$$

e será zero se, e somente se,

$$V'(t) = rV(t) + s$$

Assim, esta última equação pode ser considerada a condição necessária de otimização para a escolha do instante de venda t^* .

A interpretação econômica dessa condição recorre com facilidade ao raciocínio intuitivo: $V'(t)$ representa a taxa de variação do valor de venda, ou o incremento em V se a venda for adiada por um ano, enquanto os dois termos da direita indicam, respectivamente, os incrementos no custo do juro e o custo de armazenagem acarretado por esse adiamento da venda (receita e custo são ambos avaliados no instante t^*). Portanto, a idéia de igualar os dois lados é, para nós, apenas “um vinho velho em uma garrafa nova”, pois nada mais é do que a mesma condição $MC = MR$ com uma aparência diferente!

Valor presente de um fluxo perpétuo

Se quiséssemos que um fluxo de caixa persistisse para sempre – uma situação exemplificada pelo juro de uma obrigação perpétua ou pela receita advinda de um ativo de capital indestrutível, tal como um terreno –, o valor presente do fluxo seria

$$\Pi = \int_0^{\infty} R(t)e^{-rt} dt$$

que é uma integral imprópria.

EXEMPLO 8

Encontre o valor presente de um fluxo de renda perpétua que flui à taxa uniforme de D dólares por ano, se a taxa contínua de desconto for r . Uma vez que, ao calcular uma integral imprópria, simplesmente tomamos o limite de uma integral própria, o resultado em (14.12) ainda pode ser útil. Especificamente, podemos escrever

$$\Pi = \int_0^{\infty} D e^{-rt} dt = \lim_{y \rightarrow \infty} \int_0^y D e^{-rt} dt = \lim_{y \rightarrow \infty} \frac{D}{r} (1 - e^{-ry}) = \frac{D}{r}$$

Note que o parâmetro y (número de anos) desapareceu da resposta final. E é assim que deve ser pois, aqui, estamos tratando de um fluxo *perpétuo*. Você também pode observar que nosso resultado (valor presente = taxa de fluxo de receita ÷ taxa de desconto) corresponde exatamente à fórmula familiar para a denominada capitalização de um ativo de rendimento perpétuo.

EXERCÍCIO 14.5

1. Dadas as seguintes funções receita marginal:
 - (a) $R'(Q) = 28Q - e^{0,3Q}$
 - (b) $R'(Q) = 10(1 + Q)^{-2}$
 encontre, em cada caso, a função receita total $R(Q)$. Que condição inicial você pode introduzir para definir a constante de integração?
2. (a) Dadas a propensão marginal à importação $M'(Y) = 0,1$ e a informação que $M = 20$ quando $Y = 0$, ache a função importação $M(Y)$.
 (b) Dadas a propensão marginal ao consumo $C'(Y) = 0,8 + 0,1Y^{-1/2}$ e a informação que $C = Y$ quando $Y = 100$, ache a função consumo $C(Y)$.
3. Suponha que a taxa de investimento é descrita pela função $I(t) = 12t^{1/3}$ e que $K(0) = 25$:
 - (a) Encontre a trajetória temporal do estoque de capital K .
 - (b) Encontre o montante de acumulação de capital durante os intervalos de tempo $[0, 1]$ e $[1, 3]$, respectivamente.
4. Dado um fluxo contínuo de renda à taxa constante de \$1.000 por ano:
 - (a) Qual será o valor presente Π se o fluxo de renda durar 2 anos e a taxa contínua de desconto for 0,05 por ano?
 - (b) Qual será o valor presente Π se o fluxo de renda terminar exatamente após 3 anos e a taxa de desconto for 0,04?
5. Qual é o valor presente de um fluxo de caixa perpétuo de:
 - (a) \$1.450 por ano, descontado a $r = 5\%$?
 - (b) \$2.460 por ano, descontado a $r = 8\%$?

14.6 Modelo de crescimento de Domar

No problema de crescimento da população em (14.1) e (14.2) e no problema de formação de capital em (14.10), o objetivo comum é delinear uma trajetória temporal com base em algum padrão determinado de variação de uma variável. No clássico modelo de crescimento do professor Domar,[†] por outro lado, a idéia é estipular o tipo de trajetória temporal que deve prevalecer se uma certa condição de equilíbrio da economia tiver de ser satisfeita.

A estrutura

As premissas básicas do modelo de Domar são as seguintes:

1. Qualquer variação na taxa de fluxo de investimento por ano $I(t)$ produzirá um efeito dual: afetará a demanda agregada, bem como a capacidade produtiva da economia.
2. O efeito sobre a demanda causado por uma variação em $I(t)$ ocorre por meio do processo multiplicador que, por hipótese, funciona instantaneamente. Assim, um aumento em $I(t)$ elevará a taxa de fluxo de renda por ano $Y(t)$ por um múltiplo do incremento em $I(t)$. O multiplicador é $k = 1/s$, onde s representa a propensão marginal a poupar (constante) dada. Supondo-se que $I(t)$ é o único fluxo de dispêndio (paramétrico) que influencia a taxa de fluxo de renda, podemos declarar que

$$\frac{dY}{dt} = \frac{dI}{dt} \frac{1}{s} \quad (14.13)$$

[†] Domar, Evsey D. "Capital Expansion, Rate of Growth, and Employment". *Econometrica*, p. 137–147, abr. 1946; reimpresso em Domar, *Essays in the Theory of Economic Growth*. Fair Lawn, N.J.: Oxford University Press. 1957, p. 70–82.

3. O efeito de capacidade do investimento deve ser medido pela variação na taxa de produção *potencial* que a economia é capaz de produzir. Supondo uma razão capacidade-capital constante, podemos escrever

$$\frac{\kappa}{K} \equiv \rho \quad (= \text{uma constante})$$

onde κ (a letra grega capa) representa a capacidade ou fluxo potencial de produção por ano e ρ (a letra grega rô) denota a razão capacidade-capital dada. É claro que isso implica que, com um estoque de capital $K(t)$, a economia é potencialmente capaz de produzir um produto anual, ou renda, equivalente a $\kappa \equiv \rho K$ dólares. Note que, por $\kappa \equiv \rho K$ (a função produção), deduz-se que $d\kappa = \rho dK$, e

$$\frac{d\kappa}{dt} = \rho \frac{dK}{dt} = \rho I \quad (14.14)$$

No modelo de Domar, o equilíbrio é definido como uma situação na qual a capacidade produtiva é totalmente utilizada. Ter equilíbrio é, portanto, exigir que a demanda agregada seja exatamente igual à produção potencial possível em um ano; isto é, $Y = \kappa$. Se partirmos inicialmente de uma situação de equilíbrio, contudo, o requisito se reduzirá ao balanceamento das respectivas *variações* em capacidade e em demanda agregada, isto é,

$$\frac{dY}{dt} = \frac{d\kappa}{dt} \quad (14.15)$$

Que tipo de trajetória temporal do investimento $I(t)$ pode satisfazer essa condição de equilíbrio em todos os instantes?

Encontrando a solução

Para responder a essa pergunta, em primeiro lugar, substituímos (14.13) e (14.14) na condição de equilíbrio (14.15). O resultado é a seguinte equação diferencial:

$$\frac{dI}{dt} \frac{1}{s} = \rho I \quad \text{ou} \quad \frac{1}{I} \frac{dI}{dt} = \rho s \quad (14.16)$$

Visto que (14.16) especifica um padrão definido de variação para I , devemos ser capazes de encontrar a trajetória do investimento de equilíbrio (ou requerido) a partir da expressão (14.16).

Neste caso simples, a solução pode ser obtida integrando diretamente ambos os lados da segunda equação em (14.16) em relação a t . O fato de que os dois lados são idênticos no equilíbrio nos assegura que suas integrais são iguais. Assim,

$$\int \frac{1}{I} \frac{dI}{dt} dt = \int \rho s dt$$

Pela regra de substituição e pela regra do logaritmo, o lado esquerdo nos dá

$$\int \frac{dI}{I} = \ln |I| + c_1 \quad (I \neq 0)$$

ao passo que o lado direito resulta em (sendo ρs uma constante)

$$\int \rho s dt = \rho s t + c_2$$

Igualando os dois resultados e combinando as duas constantes, temos

$$\ln |I| = \rho s t + c \quad (14.17)$$

Para obter $|I|$ de $\ln |I|$, realizamos uma operação conhecida como “tomar o antilogaritmo de $\ln |I|$ ”, que utiliza o fato que $e^{\ln x} = x$. Assim, deixando que cada lado de (14.17) se torne o expoente da constante e , obtemos

$$e^{\ln |I|} = e^{(\rho s t + c)}$$

$$\text{ou} \quad |I| = e^{\rho s t} e^c = A e^{\rho s t} \quad \text{onde} \quad A \equiv e^c$$

Se tomarmos o investimento como positivo, então $|I| = I$, de modo que o resultado precedente se torna $I(t) = A e^{\rho s t}$, onde A é arbitrária. Para nos livrarmos dessa constante arbitrária, estabelecemos $t = 0$ na equação $I(t) = A e^{\rho s t}$, para obter $I(0) = A e^0 = A$. Isso define a constante A e nos habilita a expressar a solução – a trajetória do investimento requerido – como

$$I(t) = I(0) e^{\rho s t} \quad (14.18)$$

onde $I(0)$ denota a taxa inicial de investimento.[†]

Esse resultado tem um significado econômico inquietante. Para manter o equilíbrio entre capacidade e demanda ao longo do tempo, a taxa de fluxo de investimento deve crescer exatamente à taxa exponencial de ρs , ao longo de uma trajetória como ilustrado na Figura 14.6. Obviamente, quanto maior a razão capacidade-capital ou a propensão marginal a poupar, maior será a taxa de crescimento requerida. Mas, a qualquer taxa, uma vez conhecidos os valores de ρ e de s , a trajetória requerida de crescimento do investimento fica estabelecida com muita rigidez.

O fio da navalha

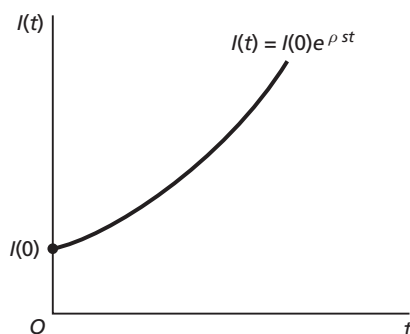
Agora fica relevante perguntar o que acontecerá se a taxa *real* de crescimento do investimento – vamos denominá-la r – for diferente da taxa *requerida* ρs .

A abordagem de Domar é definir um *coeficiente de utilização*

$$u = \lim_{t \rightarrow \infty} \frac{Y(t)}{K(t)} \quad [u = 1 \text{ significa utilização total da capacidade}]$$

e mostrar que $u = r/\rho s$, de modo que $u \geq 1$ quando $r \geq \rho s$. Em outras palavras, se houver uma discrepância entre as taxas real e requerida ($r \neq \rho s$), constataremos, no final (quando $t \rightarrow \infty$), ou uma escassez de capacidade ($u > 1$) ou um excesso de capacidade ($u < 1$), dependendo de r ser maior ou menor que ρs .

FIGURA 14.6



[†] A solução (14.18) permanecerá válida mesmo que deixemos que o investimento seja negativo no resultado $|I| = A e^{\rho s t}$. Veja Exercício 14.6-3.

Contudo, podemos mostrar que a conclusão sobre escassez e excesso de capacidade realmente se aplica em qualquer instante t , e não somente quando $t \rightarrow \infty$. Pois, uma dada taxa de crescimento r implica que

$$I(t) = I(0)e^{rt} \quad \text{e} \quad \frac{dI}{dt} = rI(0)e^{rt}$$

Portanto, por (14.13) e (14.14), temos

$$\frac{dY}{dt} = \frac{1}{s} \frac{dI}{dt} = \frac{r}{s} I(0)e^{rt}$$

$$\frac{dK}{dt} = \rho I(t) = \rho I(0)e^{rt}$$

A razão entre essas duas derivadas,

$$\frac{dY/dt}{dK/dt} = \frac{r}{\rho s}$$

deve nos informar as grandezas relativas do efeito de criação de demanda e do efeito de geração de capacidade do investimento em qualquer instante t , sob a taxa real de crescimento r . Se r (a taxa real) for maior do que ρs (a taxa requerida), então $dY/dt > dK/dt$, e o efeito da demanda ultrapassará o efeito da capacidade, causando uma escassez de capacidade. Reciprocamente, se $r < \rho s$, haverá uma deficiência na demanda agregada e, por conseguinte, um excesso de capacidade.

O curioso dessa conclusão é que, se o investimento, na realidade, crescer a uma taxa *mais veloz* que a requerida ($r > \rho s$), o resultado final será uma *escassez*, e não um excesso de capacidade. Igualmente curioso é que, se o crescimento real do investimento ficar para trás em relação à taxa requerida ($r < \rho s$), encontraremos um *excesso* de capacidade em vez de escassez. De fato, por causa desses resultados paradoxais, se permitirmos que os empreendedores ajustem a taxa real de crescimento r (até aqui considerada uma constante) de acordo com a situação de capacidade prevalecente, é praticamente certo que eles farão o tipo “errado” de ajuste. No caso de $r > \rho s$, por exemplo, a escassez de capacidade emergente motivará uma taxa de investimento ainda mais veloz. Mas isso significaria um aumento em r , em vez da redução que seria exigida sob as circunstâncias. Por consequência, a discrepância entre as duas taxas de crescimento seria intensificada, em vez de reduzida.

O resultado é que, dadas as constantes paramétricas ρ e s , o único modo de evitar a escassez, bem como o excesso de capacidade, é guiar o fluxo de investimento com muito cuidado ao longo da trajetória de equilíbrio com uma taxa de crescimento $r^* = \rho s$. E, como já mostramos, qualquer desvio em relação a esse “fio da navalha” que é a trajetória temporal acarretará uma persistente incapacidade de satisfazer a norma da utilização completa que Domar visualizou nesse modelo. E essa talvez não seja uma perspectiva muito animadora de se encarar. Felizmente, é possível conseguir resultados mais flexíveis quando são modificadas algumas premissas do modelo de Domar, como veremos no modelo de crescimento do professor Solow, que será discutido no Capítulo 15.

EXERCÍCIO 14.6

1. Quantos fatores de produção são explicitamente considerados no modelo de Domar? O que esse fato implica no que se refere à razão capital-trabalho na produção?
2. Aprendemos, na Seção 10.2, que a constante r na função exponencial Ae^{rt} representa a taxa de crescimento da função. Aplique isso a (14.16) e deduza (14.18) sem passar por integração.

3. Mostre que, mesmo que consideremos negativo o investimento na equação $I/I = Ae^{\rho st}$, ao definirmos a constante arbitrária A , ainda chegaremos à solução (14.18).
4. Mostre que o resultado em (14.18) pode ser obtido, alternativamente, encontrando – e igualando – as integrais *definidas* de ambos os lados de (14.16),

$$\frac{1}{I} \frac{dI}{dt} = \rho s$$

em relação à variável t , com limites de integração $t = 0$ e $t = t$. Lembre-se que, quando mudamos a variável de integração de t para I , os limites de integração mudarão de $t = 0$ e $t = t$, respectivamente, para $I = I(0)$ e $I = I(t)$.

CAPÍTULO 15

Tempo contínuo: equações diferenciais de primeira ordem

No modelo de crescimento de Domar, resolvemos uma equação diferencial simples por integração direta. Para equações diferenciais mais complicadas, há vários métodos estabelecidos de solução. Mesmo nestes últimos casos, entretanto, a idéia fundamental subjacente aos métodos de solução ainda são as técnicas do cálculo integral. Por essa razão, a solução de uma equação diferencial muitas vezes é denominada a *integral* daquela equação.

Neste capítulo, discutiremos apenas equações diferenciais de *primeira ordem*. Nesse contexto, a palavra *ordem* refere-se à ordem mais alta das derivadas (ou diferenciais) que aparecem na equação diferencial; assim, uma equação diferencial de primeira ordem pode conter apenas a derivada primeira, digamos, dy/dt .

15.1 Equações diferenciais lineares de primeira ordem com coeficiente constante e termo constante

A derivada primeira dy/dt é a única que pode aparecer em uma equação diferencial de primeira ordem, mas ela pode aparecer em várias potências: dy/dt , $(dy/dt)^2$ ou $(dy/dt)^3$. A potência mais alta alcançada pela derivada na equação é denominada o *grau* da equação diferencial. Caso a derivada dy/dt apareça somente no primeiro grau, o mesmo acontecendo com a variável dependente y e, além disso, não ocorrer nenhum produto da forma $y(dy/dt)$, então diz-se que a equação é *linear*. Assim, uma equação diferencial linear de primeira ordem geralmente adotará a forma[†]

$$\frac{dy}{dt} + u(t)y = w(t) \quad (15.1)$$

onde u e w são duas funções de t , assim como y . Contudo, ao contrário de dy/dt e y , não há absolutamente nenhuma restrição imposta à variável independente t . Assim, as funções u e w podem perfeitamente representar expressões como t^2 e e^t ou algumas funções mais complicadas de t ; por outro lado, u e w também podem ser constantes.

Esse último ponto nos leva a uma classificação ulterior. Quando a função u (o coeficiente da variável dependente y) for uma constante, e quando a função w for um termo aditivo constante,

[†] Note que o coeficiente do termo de derivada dy/dt em (15.1) é a unidade. Isso não quer dizer que a derivada nunca possa ter um coeficiente que não seja a unidade, mas, quando tal coeficiente aparecer, sempre podemos "normalizar" a equação dividindo cada termo pelo coeficiente citado. Por essa razão, ainda assim a forma dada em (15.1) pode ser considerada como uma representação geral.

(15.1) se reduz ao caso especial de uma equação diferencial linear de primeira ordem com *coeficiente constante e termo constante*. Nesta seção, trataremos somente dessa variedade simples de equações diferenciais.

O caso homogêneo

Se u e w forem funções constantes e se, por acaso, w for identicamente zero, (15.1) se tornará

$$\frac{dy}{dt} + ay = 0 \quad (15.2)$$

onde a é alguma constante. Diz-se que essa equação diferencial é *homogênea* em razão do termo constante zero (compare com sistemas homogêneos de equações). A característica que define uma equação homogênea é que, quando todas as variáveis (aqui, dy/dt e y) são multiplicadas por uma dada constante, a equação permanece válida. Essa característica continua valendo se o termo constante for zero, mas será perdida se o termo constante não for zero.

A equação (15.2) pode ser escrita alternativamente como

$$\frac{1}{y} \frac{dy}{dt} = -a \quad (15.2')$$

Mas você perceberá que a equação diferencial (14.16) que encontramos no modelo de Domar é exatamente dessa forma. Em conseqüência, por analogia, podemos escrever, de imediato, a solução de (15.2) ou (15.2') da seguinte maneira:

$$y(t) = Ae^{-at} \quad [\text{solução geral}] \quad (15.3)$$

$$\text{ou} \quad y(t) = y(0)e^{-at} \quad [\text{solução definida}] \quad (15.3')$$

Em (15.3), aparece uma constante arbitrária A ; por conseguinte, ela é uma *solução geral*. Quando A for substituído por qualquer valor, a solução torna-se uma *solução particular* de (15.2). Há um número infinito de soluções particulares, uma para cada valor possível de A , incluindo o valor $y(0)$. Este último valor, contudo, tem um significado especial: $y(0)$ é o único valor que pode fazer com que a solução satisfaça a condição inicial. Uma vez que isso representa o resultado de definir a constante arbitrária, vamos nos referir a (15.3') como a *solução definida* da equação diferencial (15.2) ou (15.2').

Duas coisas devem ser observadas na solução de uma equação diferencial: (1) a solução não é um valor numérico, mas uma função $y(t)$ – uma trajetória temporal se t simbolizar tempo; e (2) a solução $y(t)$ é livre de quaisquer expressões de derivada ou diferencial, de modo que, tão logo um valor específico de t seja substituído nela, um valor correspondente de y pode ser calculado diretamente.

O caso não-homogêneo

Quando uma constante diferente de zero toma o lugar do zero em (15.2), temos uma equação diferencial linear *não-homogênea*

$$\frac{dy}{dt} + ay = b \quad (15.4)$$

A solução dessa equação consistirá na soma de dois termos, um dos quais é denominado a *função complementar* (que denotaremos por y_c), e o outro é conhecido como a *solução particular* (a ser denotada por y_p). Como será demonstrado, cada um deles tem uma interpretação econômica significativa. Aqui, apresentaremos somente o método de solução; o método racional ficará claro mais adiante.

Ainda que nosso objetivo seja resolver a equação *não-homogênea* (15.4), teremos de nos referir com frequência à sua versão homogênea, como mostrado em (15.2). Por questão de conveniência, denominaremos esta última a *equação homogênea associada* a (15.4). Segundo esse critério, a equação não-homogênea (15.4) em si pode ser denominada uma *equação completa*. Acontece que a função complementar y_c nada mais é do que a solução geral da equação homogênea associada, enquanto a solução particular y_p é simplesmente *qualquer* solução particular da equação completa.

A discussão do caso homogêneo já nos deu a solução geral da equação homogênea associada e, por conseguinte, podemos escrever

$$y_c = Ae^{-at} \quad [\text{por (15.3)}]$$

E a solução particular? Uma vez que a solução particular é *qualquer* solução particular da equação completa, podemos tentar, em primeiro lugar, o tipo de solução mais simples possível, a saber, y é alguma constante ($y = k$). Se y for uma constante, então segue-se que $dy/dt = 0$ e (15.4) se tornará $ay = b$, com a solução $y = b/a$. Por conseguinte, a solução constante funcionará contanto que $a \neq 0$. Nesse caso, temos

$$y_p = \frac{b}{a} \quad (a \neq 0)$$

A soma da função complementar e da solução particular constitui, então, a solução geral da equação completa (15.4):

$$y(t) = y_c + y_p = Ae^{-at} + \frac{b}{a} \quad [\text{solução geral, caso de } a \neq 0] \quad (15.5)$$

O que torna essa solução uma solução geral é a presença da constante arbitrária A . É claro que podemos definir essa constante por meio de uma condição inicial. Digamos que y assume o valor $y(0)$ quando $t = 0$. Então, estabelecendo $t = 0$ em (15.5), constatamos que

$$y(0) = A + \frac{b}{a} \quad \text{e} \quad A = y(0) - \frac{b}{a}$$

Assim, podemos reescrever (15.5) como

$$y(t) = \left[y(0) - \frac{b}{a} \right] e^{-at} + \frac{b}{a} \quad [\text{solução definida, caso de } a \neq 0] \quad (15.5')$$

É bom observar que a utilização da condição inicial para definir a constante arbitrária é – e deve ser – executada como a etapa *final*, após termos achado a solução geral para a equação completa. Visto que os valores de ambas, y_c e y_p , estão relacionados ao valor de $y(0)$, as duas devem ser levadas em conta na definição da constante A .

EXEMPLO 1

Resolva a equação $dy/dt + 2y = 6$, com a condição inicial $y(0) = 10$. Aqui, temos $a = 2$ e $b = 6$; assim, por (15.5'), a solução é

$$y(t) = (10 - 3)e^{-2t} + 3 = 7e^{-2t} + 3$$

EXEMPLO 2

Resolva a equação $dy/dt + 4y = 0$, com a condição inicial $y(0) = 1$. Visto que $a = 4$ e $b = 0$, temos

$$y(t) = (1 - 0)e^{-4t} + 0 = e^{-4t}$$

A mesma resposta poderia ter sido obtida de (15.3'), a fórmula para o caso homogêneo. A equação homogênea (15.2) é um mero caso especial da equação não-homogênea (15.4) quando $b = 0$. Por conseqüência, a fórmula (15.3') também é um caso especial da fórmula (15.5') sob a circunstância de que $b = 0$.

E se $a = 0$, de modo que a solução em (15.5') é indefinida? Nesse caso, a equação diferencial tem a seguinte forma extremamente simples

$$\frac{dy}{dt} = b \quad (15.6)$$

Por integração direta, é possível constatar, de imediato, que sua solução geral é

$$y(t) = bt + c \quad (15.7)$$

onde c é uma constante arbitrária. Na verdade, os dois termos componentes em (15.7) podem, mais uma vez, ser identificados, respectivamente, como a função complementar e a solução particular da equação diferencial dada. Visto que $a = 0$, a função complementar pode ser expressa simplesmente como

$$y_c = Ae^{-at} = Ae^0 = A \quad (A = \text{uma constante arbitrária})$$

Quanto à solução particular, o fato de a solução constante $y = k$ não funcionar no presente caso de $a = 0$ sugere que deveríamos tentar, alternativamente, uma solução *não-constante*. Vamos considerar o tipo mais simples possível desta última, a saber, $y = kt$. Se $y = kt$, então $dy/dt = k$, e a equação completa (15.6) será reduzida a $k = b$, de modo que podemos escrever

$$y_p = bt \quad (a = 0)$$

Nossa nova solução experimental realmente funciona! A solução geral de (15.6) é, por conseguinte,

$$y(t) = y_c + y_p = A + bt \quad [\text{solução geral, caso de } a = 0] \quad (15.7')$$

cujo resultado é idêntico ao de (15.7), porque c e A nada mais são do que notações alternativas para uma constante arbitrária. Entretanto, note que, no presente caso, y_c é uma constante, enquanto y_p é uma função do tempo – o exato oposto da situação em (15.5).

Definindo a constante arbitrária, constatamos que a solução definida é

$$y(t) = y(0) + bt \quad [\text{solução definida, caso de } a = 0] \quad (15.7'')$$

EXEMPLO 3

Resolva a equação $dy/dt = 2$, com a condição inicial $y(0) = 5$. A solução é, por (15.7''),

$$y(t) = 5 + 2t$$

Verificação da solução

É verdade que a validade de todas as soluções de equações diferenciais sempre pode ser verificada por diferenciação.

Se tentarmos isso na solução (15.5'), podemos obter a derivada

$$\frac{dy}{dt} = -a \left[y(0) - \frac{b}{a} \right] e^{-at}$$

Quando essa expressão para dy/dt e a expressão para $y(t)$ mostrada em (15.5') forem substituídas no lado esquerdo da equação diferencial (15.4), aquele lado deve se reduzir exatamente ao valor do termo constante b no lado direito de (15.4) se a solução estiver correta. Fazendo essa substituição, realmente constatamos que

$$-a \left[y(0) - \frac{b}{a} \right] e^{-at} + a \left\{ \left[y(0) - \frac{b}{a} \right] e^{-at} + \frac{b}{a} \right\} = b$$

Assim, nossa solução está correta, contanto que também satisfaça a condição inicial. Para verificar esta última, vamos estabelecer $t = 0$ na solução (15.5'). Visto que o resultado

$$y(0) = \left[y(0) - \frac{b}{a} \right] + \frac{b}{a} = y(0)$$

é uma identidade, a condição inicial é, de fato, satisfeita.

Recomendamos, como uma etapa final no processo de solução de uma equação diferencial, que você adote o hábito de verificar a validade da resposta, assegurando-se que (1) a derivada da trajetória temporal $y(t)$ é consistente com a equação diferencial dada; e (2) a solução definida satisfaz a condição inicial.

EXERCÍCIO 15.1

1. Encontre y_c , y_p , a solução geral e a solução definida:

(a) $\frac{dy}{dt} + 4y = 12$; $y(0) = 2$

(c) $\frac{dy}{dt} + 10y = 15$; $y(0) = 0$

(b) $\frac{dy}{dt} - 2y = 0$; $y(0) = 9$

(d) $2\frac{dy}{dt} + 4y = 6$; $y(0) = 1\frac{1}{2}$

2. Verifique a validade de suas respostas ao Problema 1.

3. Encontre a solução de cada uma das seguintes equações utilizando uma fórmula adequada desenvolvida no texto:

(a) $\frac{dy}{dt} + y = 4$; $y(0) = 0$

(d) $\frac{dy}{dt} + 3y = 2$; $y(0) = 4$

(b) $\frac{dy}{dt} = 23$; $y(0) = 1$

(e) $\frac{dy}{dt} - 7y = 7$; $y(0) = 7$

(c) $\frac{dy}{dt} - 5y = 0$; $y(0) = 6$

(f) $3\frac{dy}{dt} + 6y = 5$; $y(0) = 0$

4. Verifique a validade de suas respostas ao Problema 3.

15.2 Dinâmica do preço de mercado

No modelo (macro) de crescimento de Domar, encontramos uma aplicação do caso *homogêneo* de equações diferenciais lineares de primeira ordem. Para ilustrar o caso *não-homogêneo*, vamos apresentar um modelo (micro) dinâmico do mercado.

A estrutura

Suponha que, para uma mercadoria específica, as funções demanda e oferta são as seguintes:

$$\begin{aligned} Q_d &= \alpha - \beta P & (\alpha, \beta > 0) \\ Q_s &= -\gamma + \delta P & (\gamma, \delta > 0) \end{aligned} \quad (15.8)$$

Então, de acordo com (3.4), o preço de equilíbrio deve ser[†]

[†] Aqui, passamos dos símbolos (a, b, c, d) de (3.4) para $(\alpha, \beta, \gamma, \delta)$ para evitar qualquer confusão possível com a utilização de a e b como parâmetros na equação diferencial (15.4), que aplicaremos, dentro em pouco, ao modelo de mercado.

$$P^* = \frac{\alpha + \gamma}{\beta + \delta} \quad (= \text{alguma constante positiva}) \quad (15.9)$$

Se acaso o preço inicial $P(0)$ estiver exatamente no nível de P^* , o mercado já estará claramente em equilíbrio e nenhuma análise dinâmica será necessária. No caso mais interessante de $P(0) \neq P^*$, contudo, P^* pode ser obtido (se for possível) somente após um devido processo de ajuste, durante o qual não somente o preço sofrerá variação ao longo do tempo, mas Q_d e Q_s , sendo funções de P , também deverão variar ao longo do tempo. Então, sob essa perspectiva, as variáveis preço e qualidade podem *todas* ser tomadas como *funções do tempo*.

Nossa pergunta dinâmica é esta: dado tempo suficiente para o processo de ajuste funcionar, ele realmente tende a trazer o preço até o nível de equilíbrio P^* ? Isto é, a trajetória temporal $P(t)$ tende a convergir para P^* , quando $t \rightarrow \infty$?

A trajetória temporal

Para responder a essa pergunta, em primeiro lugar, temos de encontrar a trajetória temporal $P(t)$. Mas isso, por sua vez, requer que, antes de mais nada, seja prescrito um padrão específico de variação de preço. Em geral, variações de preço são regidas pela intensidade relativa das forças de demanda e de oferta no mercado. Vamos supor, por questão de simplicidade, que a taxa de variação de preço (em relação ao tempo), a qualquer instante, é sempre diretamente proporcional ao *excesso de demanda* ($Q_d - Q_s$) prevalecente naquele instante. Tal padrão de variação pode ser expresso simbolicamente como

$$\frac{dP}{dt} = j (Q_d - Q_s) \quad (j > 0) \quad (15.10)$$

onde j representa um *coeficiente de ajuste* (constante). Com esse padrão de variação, podemos ter $dP/dt = 0$ se, e somente se, $Q_d = Q_s$. Nesse contexto, pode ser instrutivo observar dois sentidos para o termo *preço de equilíbrio*: o sentido intertemporal (P constante ao longo do tempo) e o sentido de compensação de mercado (o preço de equilíbrio é aquele que iguala Q_d e Q_s). No presente modelo, acontece que os dois sentidos coincidem um com o outro, mas isso pode não ser verdade para todos os modelos.

Em virtude das funções demanda e oferta em (15.8), podemos expressar (15.10) especificamente na forma

$$\frac{dP}{dt} = j (\alpha - \beta P + \gamma - \delta P) = j (\alpha + \gamma) - j (\beta + \delta) P$$

ou

$$\frac{dP}{dt} + j (\beta + \delta) P = j (\alpha + \gamma) \quad (15.10')$$

Uma vez que essa expressão está exatamente na forma da equação diferencial (15.4), e visto que o coeficiente de P é diferente de zero, podemos aplicar a fórmula de solução (15.5') e escrever a solução – a trajetória temporal do preço – como

$$\begin{aligned} P(t) &= \left[P(0) - \frac{\alpha + \gamma}{\beta + \delta} \right] e^{-j(\beta + \delta)t} + \frac{\alpha + \gamma}{\beta + \delta} \\ &= [P(0) - P^*] e^{-kt} + P^* \quad [\text{por (15.9); } k \equiv j(\beta + \delta)] \end{aligned} \quad (15.11)$$

A estabilidade dinâmica de equilíbrio

No fim, a pergunta proposta inicialmente, a saber, se $P(t) \rightarrow P^*$ quando $t \rightarrow \infty$, equivale a perguntar se o primeiro termo à direita de (15.11) tenderá a zero quando $t \rightarrow \infty$. Uma vez que $P(0)$ e

P^* são ambos constantes, o fator fundamental será a expressão exponencial e^{-kt} . Em vista do fato de que $k > 0$, aquela expressão realmente tende a zero quando $t \rightarrow \infty$. Por consequência, segundo as premissas de nosso modelo, a trajetória temporal realmente levará o preço na direção da posição de equilíbrio. Em uma situação desse tipo, na qual a trajetória temporal da variável relevante $P(t)$ converge até o nível P^* – interpretado aqui no seu papel de equilíbrio intertemporal (e não no de compensação de mercado) – diz-se que o equilíbrio é *dinamicamente estável*.

O conceito de estabilidade dinâmica é importante. Vamos examiná-lo um pouco mais, mediante uma análise mais detalhada de (15.11). Dependendo das grandezas relativas de $P(0)$ e P^* , a solução (15.11) realmente abrange três casos possíveis. O primeiro é $P(0) = P^*$, o que implica $P(t) = P^*$. Nesse evento, a trajetória temporal do preço pode ser traçada como a reta horizontal na Figura 15.1. Como mencionado anteriormente, atingir o equilíbrio, nesse caso, é um fato consumado. No segundo, podemos ter $P(0) > P^*$. Nesse caso, o primeiro termo da direita em (15.11) é positivo, mas decrescerá à medida que o aumento em t reduzir o valor de e^{-kt} . Assim, a trajetória temporal se aproximará do nível de equilíbrio P^* por cima, como ilustrado pela curva superior na Figura 15.1. No terceiro, que é o caso oposto $P(0) < P^*$, a aproximação ao nível de equilíbrio P^* será por baixo, como ilustra a curva inferior na mesma figura. Em geral, para ter estabilidade dinâmica, o *desvio* da trajetória temporal em relação ao equilíbrio deve ser ou identicamente zero (como no caso 1) ou decrescer constantemente ao longo do tempo (como nos casos 2 e 3).

Comparando (15.11) com (15.5'), vemos que o termo P^* , a contraparte de b/a , nada mais é do que a solução particular y_p , ao passo que o termo exponencial é a função complementar y_c (definida). Assim, agora temos uma interpretação econômica para y_c e y_p : y_p representa o *nível de equilíbrio intertemporal* da variável relevante e y_c é o *desvio em relação ao equilíbrio*. A estabilidade dinâmica requer a anulação assintótica da função complementar à medida que t se torna infinito.

Nesse modelo, a solução particular é uma constante, portanto temos um *equilíbrio estacionário* no sentido intertemporal, representado por P^* . Se, por outro lado, a solução particular for não-constante, como em (15.7'), podemos interpretá-lo como um *equilíbrio móvel*.

Uma utilização alternativa do modelo

O que fizemos até aqui foi analisar a estabilidade dinâmica de equilíbrio (a convergência da trajetória temporal), dadas certas especificações de sinal para os parâmetros. Um tipo alternativo de pergunta é: para assegurar a estabilidade dinâmica, quais restrições específicas devem ser impostas aos parâmetros?

A resposta a essa pergunta está contida na solução (15.11). Se permitimos que $P(0) \neq P^*$, vemos que o primeiro termo (y_c) em (15.11) tenderá a zero à medida que $t \rightarrow \infty$ se, e somente se, $k > 0$ – isto é, se, e somente se,

$$j(\beta + \delta) > 0$$

Assim, podemos considerar essa última desigualdade como a restrição necessária sobre os parâmetros j (o coeficiente de ajuste de preço), β (a negativa da inclinação da curva de demanda, desenhada com Q no eixo *vertical*) e δ (a inclinação da curva de oferta, desenhada de modo semelhante).

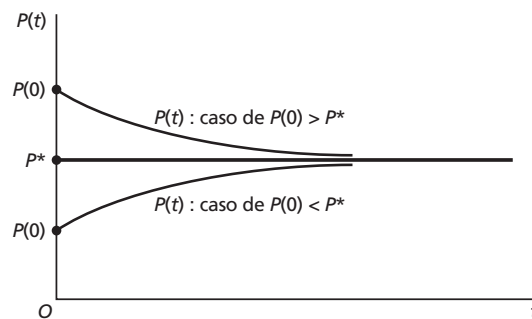


FIGURA 15.1

Caso o ajuste de preço seja do tipo “normal”, com $j > 0$, de modo que o excesso de demanda empurre o preço para cima em vez de para baixo, então essa restrição se torna simplesmente $(\beta + \delta) > 0$ ou, de modo equivalente,

$$\delta > -\beta$$

Para haver estabilidade dinâmica nesse evento, a inclinação da oferta deve exceder a inclinação da demanda. Quando ambas, demanda e oferta tiverem inclinações normais ($-\beta < 0$, $\delta > 0$), como em (15.8), esse requisito estará obviamente cumprido. Mas, mesmo que a inclinação de uma das curvas seja “perversa”, ainda assim a condição pode ser cumprida, tal como quando $\delta = 1$ e $-\beta = 1/2$ (demanda com inclinação positiva). A última situação é ilustrada na Figura 15.2, onde o preço de equilíbrio P^* é, como de hábito, determinado pelo ponto de interseção das duas curvas. Se o preço inicial por acaso estiver em P_1 , então Q_d (distância P_1G) será maior que Q_s (distância P_1F), e o excesso de demanda impulsionará os preços para cima. Por outro lado, se o preço estiver inicialmente em P_2 , então haverá um excesso de demanda *negativo* MN , que empurrará o preço para baixo. Por conseguinte, como mostram as duas setas na figura, nesse caso o ajuste de preço será *em direção ao equilíbrio*, não importando de que lado de P^* partamos. Contudo, devemos enfatizar que, conquanto essas setas possam mostrar o sentido, elas são incapazes de indicar a grandeza da variação. Assim, a Figura 15.2 é basicamente de natureza estática, e não dinâmica, e pode servir apenas para ilustrar, e não para substituir, a análise dinâmica apresentada.

EXERCÍCIO 15.2

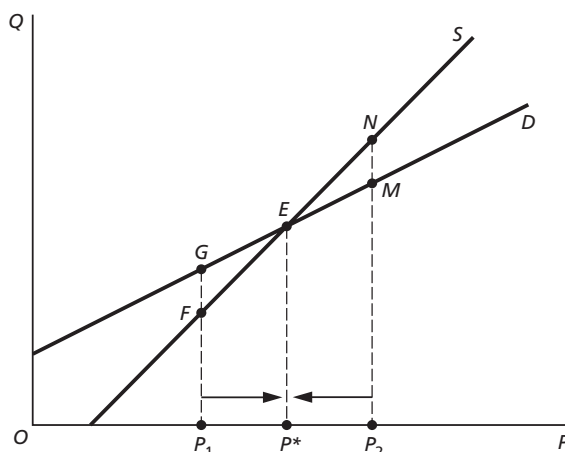
1. Se a inclinação da demanda e da oferta na Figura 15.2 forem negativas, qual curva seria mais íngreme de modo a ter estabilidade dinâmica? Sua resposta está de acordo com o critério $\delta > -\beta$?
2. Mostre que (15.10') pode ser reescrita como $dP/dt + k(P - P^*) = 0$. Se tivermos $P - P^* \equiv \Delta$ (o que significa desvio), de modo que $d\Delta/dt = dP/dt$, a equação diferencial pode ser novamente reescrita como

$$\frac{d\Delta}{dt} + k\Delta = 0$$

Encontre a trajetória temporal $\Delta(t)$, e discuta a condição para estabilidade dinâmica.

3. O modelo dinâmico de mercado discutido nesta seção é calcado no modelo estático da Seção 3.2. Qual nova característica específica é responsável pela transformação do modelo estático em dinâmico?

FIGURA 15.2



4. Sejam a demanda e a oferta

$$Q_d = \alpha - \beta P + \sigma \frac{dP}{dt} \quad Q_s = -\gamma + \delta P \quad (\alpha, \beta, \gamma, \delta > 0)$$

- Supondo que a taxa de variação do preço ao longo do tempo é diretamente proporcional ao excesso de demanda, encontre a trajetória temporal $P(t)$ (solução geral).
- Qual é o preço de equilíbrio intertemporal? Qual é o preço de equilíbrio de compensação de mercado?
- Qual restrição sobre o parâmetro σ asseguraria a estabilidade dinâmica?

5. Sejam a demanda e a oferta

$$Q_d = \alpha - \beta P + \eta \frac{dP}{dt} \quad Q_s = \delta P \quad (\alpha, \beta, \eta, \delta > 0)$$

- Supondo que o mercado seja compensado em todos os instantes de tempo, encontre a trajetória temporal $P(t)$ (solução geral).
- Esse mercado tem um preço de equilíbrio intertemporal dinamicamente estável?
- A premissa adotada no presente modelo, isto é, $Q_d = Q_s$ para todo t , é idêntica à do modelo estático de mercado na Seção 3.2. Não obstante, ainda assim temos aqui um modelo dinâmico. Como isso é possível?

15.3 Coeficiente variável e termo variável

No caso mais geral de uma equação diferencial linear de primeira ordem

$$\frac{dy}{dt} + u(t)y = w(t) \quad (15.12)$$

$u(t)$ e $w(t)$ representam um coeficiente variável e um termo variável, respectivamente. Como encontrar a trajetória temporal $y(t)$ nesse caso?

O caso homogêneo

Para o caso homogêneo, em que $w(t) = 0$, a solução ainda é fácil de obter. Visto que a equação diferencial está na forma

$$\frac{dy}{dt} + u(t)y = 0 \quad \text{ou} \quad \frac{1}{y} \frac{dy}{dt} = -u(t) \quad (15.13)$$

temos, integrando cada um dos lados por vez em relação a t ,

$$\text{Lado esquerdo} = \int \frac{1}{y} \frac{dy}{dt} dt = \int \frac{dy}{y} = \ln y + c \quad (\text{supondo } y > 0)$$

$$\text{Lado direito} = \int -u(t) dt = -\int u(t) dt$$

Neste último, o processo de integração não pode ser levado mais adiante porque não foi atribuída uma forma específica a $u(t)$; assim, temos de nos contentar com uma expressão integral, apenas. Quando igualamos os dois lados, o resultado é

$$\ln y = -c - \int u(t) dt$$

Então, a trajetória y desejada pode ser obtida tomando-se o antilogaritmo de $\ln y$:

$$y(t) = e^{\ln y} = e^{-c} e^{-\int u(t) dt} = A e^{-\int u(t) dt} \quad \text{onde } A \equiv e^{-c} \quad (15.14)$$

Essa é a solução geral da equação diferencial (15.13).

Para destacar a natureza variável do coeficiente $u(t)$, até agora temos escrito explicitamente o argumento t . Por questão de simplicidade de notação, entretanto, daqui em diante omitiremos o argumento e abreviaremos $u(t)$ para u .

Em comparação com a solução geral (15.3) para o caso de coeficiente constante, a única modificação em (15.14) é a substituição da expressão e^{-at} pela expressão mais complicada $e^{-\int u dt}$. A idéia que fundamenta essa mudança pode ser melhor entendida se interpretarmos o termo at em e^{-at} como uma integral: $\int a dt = at$ (mais uma constante que pode ser incorporada ao termo A , visto que e elevado a uma potência constante é novamente uma constante). Sob essa perspectiva, a diferença entre as duas soluções gerais, na verdade, se transforma em uma semelhança. Pois, em ambos os casos, estamos considerando o coeficiente do termo y na equação diferencial – um termo constante a em um caso, e um termo variável u no outro – e integrando-o em relação a t e depois tomando a integral resultante com sinal negativo como o expoente de e .

Uma vez obtida a solução geral, é uma questão relativamente simples obter a solução definida com a ajuda de uma condição inicial adequada.

EXEMPLO 1

Encontre a solução geral da equação $\frac{dy}{dt} + 3t^2 y = 0$. Aqui, temos $u = 3t^2$, e $\int u dt = \int 3t^2 dt = t^3 + c$.

Em consequência, por (15.14), podemos escrever a solução como

$$y(t) = A e^{-(t^3+c)} = A e^{-t^3} e^{-c} = B e^{-t^3} \quad \text{onde } B \equiv A e^{-c}$$

Observe que, se tivéssemos omitido a constante de integração c , não teríamos perdido nenhuma informação, porque teríamos obtido $y(t) = A e^{-t^3}$, que é realmente a solução idêntica, já que ambas, A e B , representam constantes arbitrárias. Em outras palavras, a expressão e^{-c} , onde a constante c faz sua única aparição, sempre pode ser incorporada pela outra constante A .

O caso não-homogêneo

Para o caso não-homogêneo, onde $w(t) \neq 0$, a solução não é tão fácil de obter. Tentaremos encontrar aquela solução via o conceito de equações diferenciais exatas, a ser discutido na Seção 15.4. Contudo, não fará mal nenhum enunciar o resultado aqui, antes: dada a equação diferencial (15.12), a solução geral é

$$y(t) = e^{-\int u dt} \left(A + \int w e^{\int u dt} dt \right) \quad (15.15)$$

onde A é uma constante arbitrária que pode ser definida se tivermos uma condição inicial adequada.

É de interesse que essa solução geral, assim como a solução no caso de coeficiente constante e termo constante, novamente consista em dois componentes aditivos. Além do mais, um desses dois, $A e^{-\int u dt}$, nada mais é do que a solução geral da equação associada homogênea, derivada anteriormente em (15.14), e tem, portanto, as características de uma função complementar.

EXEMPLO 2

Encontre a solução geral da equação $\frac{dy}{dt} + 2ty = t$. Aqui, temos

$$u = 2t \quad w = t \quad \text{e} \quad \int u dt = t^2 + k \quad (k \text{ arbitrária})$$

Assim, por (15.15), temos

$$\begin{aligned}
 y(t) &= e^{-(t^2+k)} \left(A + \int t e^{t^2+k} dt \right) \\
 &= e^{-t^2} e^{-k} \left(A + e^k \int t e^{t^2} dt \right) \\
 &= A e^{-k} e^{-t^2} + e^{-t^2} \left(\frac{1}{2} e^{t^2} + c \right) \quad [e^{-k} e^k = 1] \\
 &= (A e^{-k} + c) e^{-t^2} + \frac{1}{2} \\
 &= B e^{-t^2} + \frac{1}{2} \quad \text{onde } B \equiv A e^{-k} + c \text{ é arbitrária.}
 \end{aligned}$$

Mais uma vez, a validade dessa solução pode ser verificada por diferenciação.

É interessante notar que, neste exemplo, poderíamos novamente ter omitido a constante de integração k , bem como a constante de integração c , sem afetar o resultado final. Isso porque ambas, k e c , podem ser incorporadas à constante arbitrária B na solução final. Aconselhamos que você experimente o processo mais simples de aplicar (15.15) sem usar as constantes k e c , e verifique se obtém a mesma solução.

EXEMPLO 3 Resolva a equação $\frac{dy}{dt} + 4ty = 4t$. Desta vez, omitiremos as constantes de integração. Visto que

$$u = 4t \quad w = 4t \quad \text{e} \quad \int u dt = 2t^2 \quad [\text{constante omitida}]$$

a solução geral é, por (15.15),

$$\begin{aligned}
 y(t) &= e^{-2t^2} \left(A + \int 4t e^{2t^2} dt \right) = e^{-2t^2} (A + e^{2t^2}) \quad [\text{constante omitida}] \\
 &= A e^{-2t^2} + 1
 \end{aligned}$$

Como era de se esperar, a omissão das constantes de integração serve para simplificar substancialmente o procedimento.

A equação diferencial $\frac{dy}{dt} + uy = w$ em (15.12) é mais geral do que a equação $\frac{dy}{dt} + ay = b$ em (15.4), visto que u e w não são necessariamente constantes, como são a e b . De acordo com isso, a fórmula de solução (15.15) também é mais geral do que a fórmula de solução (15.5). Na verdade, quando estabelecemos $u = a$ e $w = b$, (15.15) deveria se reduzir a (15.5). E é isso mesmo que acontece pois, quando temos

$$u = a \quad w = b \quad \text{e} \quad \int u dt = at \quad [\text{constante omitida}],$$

então (15.15) se torna

$$\begin{aligned}
 y(t) &= e^{-at} \left(A + \int b e^{at} dt \right) = e^{-at} \left(A + \frac{b}{a} e^{at} \right) \quad [\text{constante omitida}] \\
 &= A e^{-at} + \frac{b}{a}
 \end{aligned}$$

que é idêntica a (15.5).

EXERCÍCIO 15.3

Resolva as seguintes equações diferenciais lineares de primeira ordem; se for dada uma condição inicial, defina a constante arbitrária:

1. $\frac{dy}{dt} + 5y = 15$
2. $\frac{dy}{dt} + 2ty = 0$

3. $\frac{dy}{dt} + 2ty = t; y(0) = \frac{3}{2}$
4. $\frac{dy}{dt} + t^2 y = 5t^2; y(0) = 6$
5. $2\frac{dy}{dt} + 12y + 2e^t = 0; y(0) = \frac{6}{7}$
6. $\frac{dy}{dt} + y = t$

15.4 Equações diferenciais exatas

Agora introduziremos o conceito de equações diferenciais exatas e usaremos o método de solução pertinente para obter a fórmula de solução (15.15) citada anteriormente para a equação diferencial (15.12). Embora nosso propósito imediato seja usá-la para resolver uma equação diferencial *linear*, uma equação diferencial pode ser, por si, linear ou não-linear.

Equações diferenciais exatas

Dada uma função de duas variáveis $F(y, t)$, sua diferencial total é

$$dF(y, t) = \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial t} dt$$

Quando essa diferencial é igualada a zero, a equação resultante,

$$\frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial t} dt = 0$$

é conhecida como uma *equação diferencial exata*, porque seu lado esquerdo é exatamente a diferencial da função $F(y, t)$. Por exemplo, dada

$$F(y, t) = y^2 t + k \quad (k \text{ uma constante})$$

a diferencial total é

$$dF = 2yt \, dy + y^2 \, dt$$

assim, a equação diferencial

$$2yt \, dy + y^2 \, dt = 0 \quad \text{ou} \quad \frac{dy}{dt} + \frac{y^2}{2yt} = 0 \quad (15.16)$$

é exata.

Em geral, uma equação diferencial

$$M \, dy + N \, dt = 0 \quad (15.17)$$

é exata se, e somente se, existir uma função $F(y, t)$ tal que $M = \partial F / \partial y$ e $N = \partial F / \partial t$. Contudo, pelo teorema de Young, que afirma que $\partial^2 F / \partial t \partial y = \partial^2 F / \partial y \partial t$, também podemos afirmar que (15.17) é exata se, e somente se,

$$\frac{\partial M}{\partial t} = \frac{\partial N}{\partial y} \quad (15.18)$$

Esta última equação nos dá um teste simples para a exatidão de uma equação diferencial. Aplicando a (15.16), onde $M = 2yt$ e $N = y^2$, esse teste resulta em $\partial M/\partial t = 2y = \partial N/\partial y$; assim, a exatidão da equação diferencial citada é devidamente verificada.

Note que nenhuma restrição foi imposta aos termos M e N no que se refere ao modo pelo qual a variável y ocorre. Assim, uma equação diferencial exata pode muito bem ser *não-linear* (em y). Não obstante, ela sempre será de primeira ordem e de primeiro grau.

Como é exata, a equação diferencial diz apenas que

$$dF(y, t) = 0$$

Assim, sua solução geral deve estar claramente na forma

$$F(y, t) = c$$

Portanto, resolver uma equação diferencial exata é, basicamente, procurar a função (primitiva) $F(y, t)$ e então igualar essa função a uma constante arbitrária. Vamos delinear um método para fazer isso para a equação $M dy + N dt = 0$.

Método de solução

Para começar, uma vez que $M = \partial F/\partial y$, a função F deve conter a integral de M em relação à variável y ; então, podemos escrever um resultado preliminar – de uma forma ainda indeterminada – como se segue:

$$F(y, t) = \int M dy + \psi(t) \quad (15.19)$$

Aqui, M , uma derivada *parcial*, deve ser integrada em relação a y apenas; isto é, t deve ser tratada como uma constante no processo de integração, exatamente como foi tratada como uma constante na diferenciação parcial de $F(y, t)$, que resultou em $M = \partial F/\partial y$.[†] Uma vez que, ao diferenciar $F(y, t)$ parcialmente em relação a y , qualquer termo aditivo que contenha somente a variável t e/ou algumas constantes (mas nenhum y) seria descartado, agora devemos tomar cuidado para reintegrar esses termos no processo de integração. Isso explica por que introduzimos um termo geral, $\psi(t)$, em (15.19), o qual, embora não seja exatamente o mesmo que uma constante de integração, desempenha um papel exatamente idêntico ao da última. É relativamente fácil obter $\int M dy$; mas como determinamos a forma exata desse termo $\psi(t)$?

O truque é utilizar o fato de que $N = \partial F/\partial t$. Mas o procedimento é melhor explicado com a ajuda de exemplos específicos.

EXEMPLO 1 Resolva a equação diferencial exata

$$2yt dy + y^2 dt = 0 \quad [\text{reproduzida de (15.16)}]$$

Nessa equação, temos

$$M = 2yt \quad \text{e} \quad N = y^2$$

ETAPA i Por (15.19), podemos escrever primeiro o resultado preliminar

$$F(y, t) = \int 2yt dy + \psi(t) = y^2 t + \psi(t)$$

Note que omitimos a constante de integração porque ela pode ser automaticamente incorporada à expressão $\psi(t)$.

[†] Alguns autores empregam o símbolo de operador $\int (\cdot \cdot \cdot) \partial y$ para enfatizar que a integração é em relação a y somente. Aqui, ainda utilizaremos o símbolo $\int (\cdot \cdot \cdot) dy$, visto que há pouca possibilidade de confusão.

ETAPA ii Se diferenciarmos o resultado da Etapa i parcialmente em relação a t , podemos obter

$$\frac{\partial F}{\partial t} = y^2 + \psi'(t)$$

Mas, visto que $N = \partial F / \partial t$, podemos igualar $N = y^2$ e $\partial F / \partial t = y^2 + \psi'(t)$, para obter

$$\psi'(t) = 0$$

ETAPA iii A integração do último resultado nos dá

$$\psi(t) = \int \psi'(t) dt = \int 0 dt = k$$

e agora temos uma forma específica de $\psi(t)$. Acontece que, no presente caso, esse $\psi(t)$ é simplesmente uma constante; em geral, ele pode ser uma função não-constante de t .

ETAPA iv Os resultados das Etapas i e iii podem ser combinados para resultar

$$F(y, t) = y^2 t + k$$

Então, a solução da equação diferencial exata deve ser $F(y, t) = c$. Mas, visto que a constante k pode ser incorporada a c , podemos escrever a solução simplesmente como

$$y^2 t = c \quad \text{ou} \quad y(t) = ct^{-1/2}$$

onde c é arbitrária.

EXEMPLO 2

Resolva a equação $(t + 2y) dy + (y + 3t^2) dt = 0$. Em primeiro lugar, vamos verificar se essa é uma equação diferencial exata. Estabelecendo $M = t + 2y$ e $N = y + 3t^2$, constatamos que $\partial M / \partial t = 1 = \partial N / \partial y$. Assim, a equação passa no teste de exatidão. Para achar sua solução, seguimos novamente o procedimento delineado no Exemplo 1.

ETAPA i Aplique (15.19) e escreva

$$F(y, t) = \int (t + 2y) dy + \psi(t) = yt + y^2 + \psi(t) \quad [\text{constante incorporada a } \psi(t)]$$

ETAPA ii Diferencie esse resultado em relação a t , para obter

$$\frac{\partial F}{\partial t} = y + \psi'(t)$$

Então, igualando isso a $N = y + 3t^2$, constatamos que

$$\psi'(t) = 3t^2$$

ETAPA iii Integre este último resultado para obter

$$\psi(t) = \int 3t^2 dt = t^3 \quad [\text{constante pode ser omitida}]$$

ETAPA iv Combine os resultados das Etapas i e iii para obter a forma completa da função $F(y, t)$:

$$F(y, t) = yt + y^2 + t^3$$

o que implica que a solução da equação diferencial dada é

$$yt + y^2 + t^3 = c$$

Você deve verificar que igualar a diferencial total dessa equação a zero de fato produzirá a equação diferencial dada.

Esse procedimento em quatro etapas pode ser utilizado para resolver qualquer equação diferencial exata. O interessante é que ele pode ser aplicável até mesmo quando a equação dada *não* é exata. Contudo, para verificar isso, precisamos antes introduzir o conceito de fator *integrante*.

Fator de integrante

Às vezes uma equação diferencial inexacta pode ser transformada em exata multiplicando-se cada um de seus termos por um determinado fator comum. Tal fator é denominado *fator integrante*.

EXEMPLO 3 A equação diferencial

$$2t \, dy + y \, dt = 0$$

não é exata porque não satisfaz (15.18):

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial t}(2t) = 2 \neq \frac{\partial N}{\partial y} = \frac{\partial}{\partial y}(y) = 1$$

Contudo, se multiplicarmos cada termo por y , a equação dada se transformará em (15.16), que já determinamos ser exata. Assim, y é um fator integrante para a equação diferencial no presente exemplo.

Quando é possível achar um fator integrante para uma equação diferencial inexacta, sempre é possível convertê-la em exata e, então, o procedimento de solução em quatro etapas pode ser posto imediatamente em uso.

Solução de equações diferenciais lineares de primeira ordem

A equação diferencial linear geral de primeira ordem

$$\frac{dy}{dt} + uy = w$$

a qual, no formato de (15.17), pode ser expressa como

$$dy + (uy - w) \, dt = 0 \quad (15.20)$$

tem o fator de integrante

$$e^{\int u \, dt} \equiv \exp\left(\int u \, dt\right)$$

Esse fator integrante, cuja forma não é, de modo algum, intuitivamente óbvia, pode ser “descoberto” como se segue. Seja I o fator integrante (ainda desconhecido). A multiplicação de todos os termos de (15.20) por I deve convertê-la em uma equação diferencial exata

$$\underbrace{I}_{\tilde{M}} dy + \underbrace{I(uy - w)}_{\tilde{N}} dt = 0 \quad (15.20')$$

O teste de exatidão determina que $\partial \tilde{M} / \partial t = \partial \tilde{N} / \partial y$. A inspeção visual das expressões $\tilde{M}$ e $\tilde{N}$ sugere que, uma vez que $\tilde{M}$ consiste em I somente, e visto que u e w são funções apenas de t , o teste de exatidão será reduzido a uma condição muito simples se I também for uma função apenas de t . Pois, então, o teste $\partial \tilde{M} / \partial t = \partial \tilde{N} / \partial y$ se torna

$$\frac{dI}{dt} = Iu \quad \text{ou} \quad \frac{dI/dt}{I} = u$$

Assim, a forma especial $I = I(t)$ realmente pode funcionar, contanto que tenha uma taxa de crescimento igual a u , ou, mais explicitamente, a $u(t)$. De acordo com isso, $I(t)$ deve assumir a forma específica

$$I(t) = Ae^{\int u dt} \quad [\text{conforme (15.13) e (15.14)}]$$

Entretanto, como podemos verificar com facilidade, a constante A pode ser igualada a 1 sem afetar a capacidade de $I(t)$ de cumprir o teste de exatidão. Assim, podemos usar a forma mais simples $e^{\int u dt}$ como o fator integrante.

Substituindo esse fator integrante em (15.20'), temos a equação diferencial exata

$$e^{\int u dt} dy + e^{\int u dt} (uy - w) dt = 0 \quad (15.20'')$$

que então pode ser resolvida pelo procedimento em quatro etapas.

ETAPA i Primeiro, aplicamos (15.19) para obter

$$F(y, t) = \int e^{\int u dt} dy + \psi(t) = ye^{\int u dt} + \psi(t)$$

O resultado da integração surge nessa forma simples porque o integrando é independente da variável y .

ETAPA ii Em seguida, diferenciamos o resultado da Etapa i em relação a t para obter

$$\frac{\partial F}{\partial t} = ye^{\int u dt} + \psi'(t) \quad [\text{regra da cadeia}]$$

E, uma vez que essa expressão pode ser igualada a $N = e^{\int u dt}(uy - w)$, temos

$$\psi'(t) = -we^{\int u dt}$$

ETAPA iii A integração direta agora resulta em

$$\psi(t) = - \int we^{\int u dt} dt$$

Considerando que não foram atribuídas formas definidas às funções $u = u(t)$ e $w = w(t)$, nada mais pode ser feito com essa integral, e temos de nos contentar com essa expressão bastante geral para $\psi(t)$.

ETAPA iv Substituindo essa expressão $\psi(t)$ no resultado da Etapa i, constatamos que

$$F(y, t) = ye^{\int u dt} - \int we^{\int u dt} dt$$

Assim, a solução geral da equação diferencial exata (15.20'') – e da equação diferencial linear de primeira ordem equivalente, porém inexata, (15.20) – é

$$ye^{\int u dt} - \int we^{\int u dt} dt = c$$

Após rearranjar e substituir o símbolo c (da constante arbitrária) por A , essa expressão pode ser escrita como

$$y(t) = e^{-\int u dt} \left(A + \int we^{\int u dt} dt \right) \quad (15.21)$$

que é exatamente o resultado dado anteriormente em (15.15).

EXERCÍCIO 15.4

- Verifique se cada uma das seguintes equações diferenciais é exata e resolva pelo procedimento em quatro etapas:
 - $2yt^3 dy + 3y^2t^2 dt = 0$
 - $3y^2t dy + (y^3 + 2t) dt = 0$
 - $t(1 + 2y) dy + y(1 + y) dt = 0$
 - $\frac{dy}{dt} + \frac{2y^4t + 3t^2}{4y^3t^2} = 0$ [Sugestão: Primeiro, converta-a à forma (15.17).]
- As seguintes equações diferenciais são exatas? Se não forem, experimente t , y e y^2 como possíveis fatores integrantes.
 - $2(t^3 + 1) dy + 3yt^2 dt = 0$
 - $4y^3t dy + (2y^4 + 3t) dt = 0$
- Aplicando o procedimento em quatro etapas à equação diferencial exata geral $M dy + N dt = 0$, deduza a seguinte fórmula para a solução geral de uma equação diferencial exata:

$$\int M dy + \int N dt - \int \left(\frac{\partial}{\partial t} \int M dy \right) dt = c$$

15.5 Equações diferenciais não-lineares de primeira ordem e de primeiro grau

Em uma equação diferencial *linear*, restringimos ao *primeiro grau* não somente a derivada dy/dt , mas também a variável dependente y , e não permitimos que o produto $y(dy/dt)$ apareça. Quando y aparece com uma potência maior que um, a equação torna-se *não-linear* mesmo que contenha somente a derivada dy/dt no primeiro grau. Em geral, uma equação na forma

$$f(y, t) dy + g(y, t) dt = 0 \quad (15.22)$$

ou

$$\frac{dy}{dt} = h(y, t) \quad (15.22')$$

onde não há nenhuma restrição às potências de y e t , constitui uma equação diferencial não-linear de primeira ordem de primeiro grau porque dy/dt é uma derivada de primeira ordem na primeira potência. Certas variedades dessas equações podem ser resolvidas com relativa facilidade por procedimentos mais ou menos rotineiros. Discutiremos brevemente três casos.

Equações diferenciais exatas

O primeiro é o caso – agora familiar – das equações diferenciais exatas. Como salientamos anteriormente, a variável y pode aparecer em uma equação exata com uma potência alta, como em (15.16) – $2yt dy + y^2 dt = 0$ –, que você deve comparar com (15.22). É verdade que o cancelamento do fator comum y de ambos os termos da esquerda reduzirá a equação a uma forma linear, mas a propriedade de exatidão será perdida, nesse caso. Logo, por ser uma equação diferencial *exata*, ela deve ser considerada não-linear.

Uma vez que o método de solução para equações diferenciais exatas já foi discutido, nenhum outro comentário adicional é necessário.

Variáveis separáveis

Pode acontecer de a equação diferencial em (15.22)

$$f(y, t) dy + g(y, t) dt = 0$$

possuir a conveniente propriedade de a função f depender apenas da variável y , enquanto a função g envolve somente a variável t , de modo que a equação se reduz à forma especial

$$f(y) dy + g(t) dt = 0 \quad (15.23)$$

Nesse caso, diz-se que as variáveis são *separáveis*, porque os termos que envolvem y – consolidados em $f(y)$ – podem ser matematicamente separados dos termos que envolvem t , que são agrupados sob $g(t)$. Para resolver esse tipo especial de equação são necessárias apenas técnicas simples de integração.

EXEMPLO 1 Resolva a equação $3y^2 dy - t dt = 0$. Primeiro, vamos reescrever a equação como

$$3y^2 dy = t dt$$

Integrando os dois lados (cada um dos quais é uma diferencial), e igualando os resultados, obtemos

$$\int 3y^2 dy = \int t dt \quad \text{ou} \quad y^3 + c_1 = \frac{1}{2} t^2 + c_2$$

Assim, a solução geral pode ser escrita como

$$y^3 = \frac{1}{2} t^2 + c_2 \quad \text{ou} \quad y(t) = \left(\frac{1}{2} t^2 + c \right)^{1/3}$$

O ponto notável aqui é que a integração de cada termo é executada em relação a uma variável diferente; é isso que torna a equação de variáveis separáveis comparativamente fácil de manusear.

EXEMPLO 2 Resolva a equação $2t dy + y dt = 0$. À primeira vista, essa equação diferencial não parece pertencer a esse assunto, porque ela não está de acordo com a forma geral de (15.23). Para sermos específicos, consideramos que os coeficientes de dy e dt envolvem as variáveis “erradas”. Todavia, uma simples transformação – dividir tudo por $2yt$ ($\neq 0$) – reduzirá a equação à forma de variável separável

$$\frac{1}{y} dy + \frac{1}{2t} dt = 0$$

De nossa experiência com o Exemplo 1, podemos avançar em direção à solução (sem antes transpor um termo) como se segue:[†]

$$\int \frac{1}{y} dy + \int \frac{1}{2t} dt = c$$

portanto,

$$\ln y + \frac{1}{2} \ln t = c \quad \text{ou} \quad \ln(yt^{1/2}) = c$$

Assim, a solução é

$$yt^{1/2} = e^c = k \quad \text{ou} \quad y(t) = kt^{-1/2}$$

onde k é uma constante arbitrária, assim como os símbolos c e A utilizados em outros lugares.

[†] No resultado da integração, deveríamos ter escrito, em termos estritos, $\ln |y|$ e $\frac{1}{2} \ln |t|$. Se pudermos supor que y e t são positivos, como é adequado na maioria dos contextos econômicos, então ocorrerá o resultado dado no texto.

Note que, em vez de resolver a equação no Exemplo 2 como fizemos, também poderíamos tê-la transformado antes em uma equação diferencial exata (usando o fator integrante y) e então tê-la resolvido como tal. É claro que a solução, já dada no Exemplo 1 da Seção 15.4, deve ser idêntica àquela que acabamos de obter por separação de variáveis. O importante é que uma equação diferencial dada muitas vezes pode ser solucionada de mais de uma maneira e, portanto, podemos escolher o método a ser utilizado. Em outros casos, uma equação diferencial que não se presta a um método específico pode, ainda assim, vir a se prestar após uma transformação apropriada.

Equações redutíveis à forma linear

Se a equação diferencial $dy/dt = h(y, t)$ acaso assumir a forma não-linear específica

$$\frac{dy}{dt} + Ry = Ty^m \quad (15.24)$$

onde R e T são duas funções de t , e m é qualquer número, exceto 0 e 1 (e se $m = 0$ ou $m = 1$?), então a equação – denominada uma *equação de Bernoulli* – sempre pode ser reduzida a uma equação diferencial linear e resolvida como tal.

O procedimento de redução é relativamente simples. Em primeiro lugar, podemos dividir (15.24) por y^m , para obter

$$y^{-m} \frac{dy}{dt} + Ry^{1-m} = T$$

Se introduzirmos uma nova variável z , como se segue:

$$z = y^{1-m} \quad \left[\text{de modo que } \frac{dz}{dt} = \frac{dz}{dy} \frac{dy}{dt} = (1-m)y^{-m} \frac{dy}{dt} \right]$$

então a equação precedente pode ser escrita como

$$\frac{1}{1-m} \frac{dz}{dt} + Rz = T$$

Além disso, após multiplicar tudo por $(1-m) dt$ e rearranjar, podemos transformar a equação em

$$dz + [(1-m)Rz - (1-m)T] dt = 0 \quad (15.24')$$

Essa expressão é vista como uma equação diferencial linear de primeira ordem da forma (15.20), na qual a variável z tomou o lugar de y .

É claro que podemos aplicar a fórmula (15.21) para encontrar sua solução $z(t)$. Então, como etapa final, podemos traduzir z de volta para y por substituição inversa.

EXEMPLO 3

Resolva a equação $dy/dt + ty = 3ty^2$. Essa é uma equação de Bernoulli, com $m = 2$ (que nos dá $z = y^{1-m} = y^{-1}$), $R = t$, e $T = 3t$. Assim, por (15.24'), podemos escrever a equação diferencial linearizada como

$$dz + (-tz + 3t) dt = 0$$

Aplicando a fórmula (15.21), podemos constatar que a solução é

$$z(t) = A \exp\left(\frac{1}{2}t^2\right) + 3$$

(Como exercício, determine o caminho que leva a essa solução).

Uma vez que nosso interesse primário está na solução $y(t)$, e não em $z(t)$, devemos realizar uma transformação inversa usando a equação $z = y^{-1}$, ou $y = z^{-1}$. Tomando a recíproca de $z(t)$, portanto, obtemos

$$y(t) = \frac{1}{A \exp\left(\frac{1}{2}t^2\right) + 3}$$

como a solução desejada. Essa é uma solução geral, porque está presente uma constante arbitrária A .

EXEMPLO 4 Resolva a equação $dy/dt + (1/t)y = y^3$. Aqui, temos $m = 3$ (assim, $z = y^{-2}$), $R = 1/t$, e $T = 1$; assim, a equação pode ser linearizada para a forma

$$dz + \left(\frac{-2}{t}z + 2\right)dt = 0$$

Como você pode verificar, pela utilização da fórmula (15.21), a solução dessa equação diferencial é

$$z(t) = At^2 + 2t$$

Então, pela transformação inversa $y = z^{-1/2}$, segue-se que a solução geral na variável original deve ser escrita como

$$y(t) = (At^2 + 2t)^{-1/2}$$

Como exercício, verifique, por diferenciação, a validade das soluções desses dois últimos exemplos.

EXERCÍCIO 15.5

- Determine, para cada uma das seguintes: (1) se as variáveis são separáveis; e (2) se a equação é linear ou se pode ser linearizada:

(a) $2t \, dy + 2y \, dt = 0$

(c) $\frac{dy}{dt} = -\frac{t}{y}$

(b) $\frac{y}{y+t} \, dy + \frac{2t}{y+t} \, dt = 0$

(d) $\frac{dy}{dt} = 3y^2t$

- Resolva (a) e (b) no Problema 1 por separação de variáveis, considerando y e t como positivos. Verifique suas respostas por diferenciação.
- Resolva (c) no Problema 1 como uma equação de variáveis separáveis e, também, como uma equação de Bernoulli.
- Resolva (d) no Problema 1 como uma equação de variáveis separáveis e, também, como uma equação de Bernoulli.
- Verifique a correção da solução intermediária $z(t) = At^2 + 2t$ no Exemplo 4 mostrando que sua derivada dz/dt é consistente com a equação diferencial linearizada.

15.6 A abordagem gráfico-qualitativa

Os diversos casos de equações diferenciais não-lineares previamente discutidos (equações diferenciais exatas, equações de variáveis separáveis e equações de Bernoulli) foram todos resolvidos *quantitativamente*. Isto é, em cada caso procuramos e encontramos uma trajetória temporal $y(t)$ que, para cada valor de t , nos dá o valor específico correspondente da variável y .

Às vezes, é possível que não consigamos encontrar uma solução quantitativa para uma dada equação diferencial. Nesses casos, ainda assim pode ser possível averiguar as propriedades *qualitativas* da trajetória temporal – principalmente, se $y(t)$ converge – observando diretamente a própria equação diferencial ou analisando seu gráfico. Além do mais, mesmo quando há soluções quantitativas disponíveis, ainda podemos empregar as técnicas de análise qualitativa se o aspecto qualitativo da trajetória temporal for nossa preocupação principal ou exclusiva.

O diagrama de fase

Dada uma equação diferencial de primeira ordem na forma geral

$$\frac{dy}{dt} = f(y)$$

linear ou não-linear na variável y , podemos desenhar o gráfico dy/dt em função de y como na Figura 15.3. Essa representação geométrica, viável sempre que dy/dt for uma função apenas de y , é denominada um *diagrama de fase*, e o gráfico que representa a função f , uma *linha de fase*. (Uma equação diferencial dessa forma – na qual a variável t não aparece como um argumento isolado da função f – é denominada uma equação diferencial *autônoma*.) Uma vez conhecida uma linha de fase, sua configuração comunicará informações qualitativas significativas a respeito da trajetória temporal $y(t)$. A pista é dada pelas duas observações gerais a seguir:

1. Em qualquer lugar *acima* do eixo horizontal (onde $dy/dt > 0$), y tem de estar crescendo ao longo do tempo e, no que concerne ao eixo y , deve estar se movimentando da esquerda para a direita. Por um raciocínio análogo, qualquer ponto *abaixo* do eixo horizontal tem de estar associado a um movimento para a esquerda na variável y , porque a negatividade de dy/dt significa que y decresce com o tempo. Essas tendências direcionais explicam por que as pontas das setas que ilustram as linhas de fase na Figura 15.3 estão desenhadas daquela maneira. Acima do eixo horizontal, as setas apontam uniformemente para a direita – na direção nordeste ou sudeste, conforme o caso. O oposto vale abaixo do eixo y . Além disso, esses resultados são independentes do sinal algébrico de y ; mesmo que a linha de fase A (ou qualquer outra) seja transplantada para a esquerda do eixo vertical, a direção das setas não será afetada.
2. Um nível de equilíbrio de y – no sentido intertemporal do termo – se existir, pode ocorrer apenas no eixo horizontal, onde $dy/dt = 0$ (y estacionário ao longo do tempo). Por conseguinte, para encontrar um equilíbrio, basta considerar a interseção da linha de fase com o eixo y .[†] Por outro lado, para testar a estabilidade dinâmica de equilíbrio, devemos também verificar se, independentemente da posição inicial de y , a linha de fase sempre o guiará em direção à posição de equilíbrio na interseção citada.

Tipos de trajetória temporal

Com base nas observações gerais precedentes, podemos observar três tipos diferentes de trajetória temporal pelas linhas de fase ilustrativas na Figura 15.3.

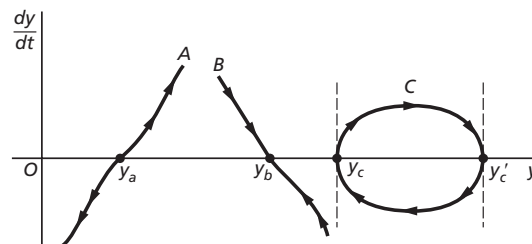


FIGURA 15.3

[†] Todavia, nem todas as interseções representam posições de equilíbrio. Veremos isso quando discutirmos a linha de fase C na Figura 15.3.

A linha de fase A tem um equilíbrio no ponto y_a ; mas tanto *acima* quanto *abaixo* daquele ponto as pontas das setas se afastam consistentemente do equilíbrio. Assim, embora o equilíbrio possa ser atingido caso $y(0) = y_a$, a situação mais usual em que $y(0) \neq y_a$ resultará em y sempre crescente [se $y(0) > y_a$] ou sempre decrescente [se $y(0) < y_a$]. Além do mais, nesse caso, o desvio de y em relação a y_a tende a crescer a um passo crescente porque, à medida que seguimos as pontas das setas da linha de fase, nos desviamos cada vez mais do eixo y e, por isso, encontramos valores numéricos de dy/dt também sempre crescentes. Por conseguinte, a trajetória temporal $y(t)$ indicada pela linha de fase A pode ser representada pelas curvas mostradas na Figura 15.4a, onde aparece o gráfico de y em função de t (e não de dy/dt em função de y). O equilíbrio y_a é dinamicamente instável.

Ao contrário, a linha de fase B implica um equilíbrio estável em y_b . Se $y(0) = y_b$, o equilíbrio prevalece imediatamente. Mas, o aspecto importante da linha de fase B é que, mesmo quando $y(0) \neq y_b$, o movimento ao longo da linha de fase guiará y em direção ao nível de y_b . A trajetória temporal $y(t)$ correspondente a esse tipo de linha de fase deve ser, portanto, da forma mostrada na Figura 15.4b, que nos lembra o modelo dinâmico de mercado.

A discussão precedente sugere que, em geral, a inclinação da linha de fase em sua interseção é a chave da estabilidade dinâmica de equilíbrio ou da convergência da trajetória temporal. Uma inclinação *positiva* (finita), tal como no ponto y_a , quer dizer *instabilidade* dinâmica; ao passo que uma inclinação *negativa* (finita), tal como em y_b , implica *estabilidade* dinâmica.

Essa generalização pode nos ajudar a extrair inferências qualitativas sobre equações diferenciais dadas sem nem mesmo fazer o gráfico de suas linhas de fase. Considere a equação diferencial linear em (15.4), por exemplo:

$$\frac{dy}{dt} + ay = b \quad \text{ou} \quad \frac{dy}{dt} = -ay + b$$

Visto que a linha de fase terá, obviamente, a inclinação (constante) $-a$ que, aqui, supomos diferente de zero, podemos inferir imediatamente (sem desenhar a linha) que

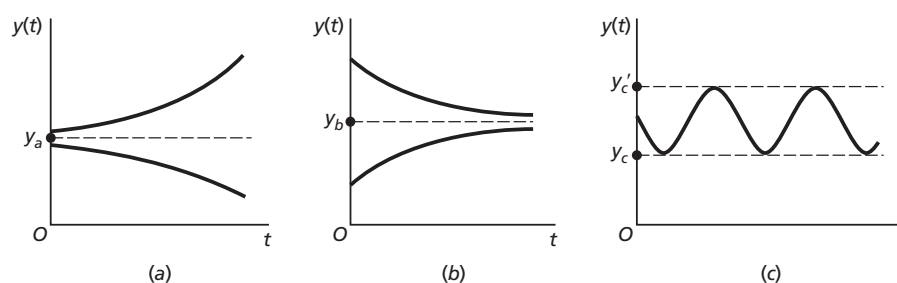
$$a \geq 0 \quad \Leftrightarrow \quad y(t) \begin{cases} \text{converge para} \\ \text{diverge do} \end{cases} \text{equilíbrio}$$

Como era de se esperar, esse resultado coincide perfeitamente com o que a solução quantitativa dessa equação nos diz:

$$y(t) = \left[y(0) - \frac{b}{a} \right] e^{-at} + \frac{b}{a} \quad [\text{de (15.5')}]$$

Aprendemos que, partindo de uma posição inicial de não-equilíbrio, a convergência de $y(t)$ depende de $e^{-at} \rightarrow 0$ quando $t \rightarrow \infty$. Isso pode acontecer se, e somente se, $a > 0$; se $a < 0$, então $e^{-at} \rightarrow \infty$ quando $t \rightarrow \infty$, e $y(t)$ não pode convergir. Assim, nossa conclusão é uma só e a mesma, quer cheguemos a ela pela via quantitativa, quer pela via qualitativa.

FIGURA 15.4



Resta discutir a linha de fase C , a qual, por ser um laço fechado ao longo do eixo horizontal, não se qualifica como uma *função*, mas, por sua vez, mostra uma *relação* entre dy/dt e y .[†] O elemento novo e interessante que surge nesse caso é a possibilidade de uma trajetória temporal que oscila periodicamente. Do modo como a linha C está desenhada, veremos que y oscila entre os dois valores y_c e y'_c em um movimento perpétuo. Para gerar a oscilação periódica, é claro que o laço deve abarcar o eixo horizontal de tal maneira que dy/dt possa ser alternadamente positiva e negativa. Além disso, nos dois pontos de interseção y_c e y'_c , a linha de fase deve ter uma inclinação infinita; caso contrário, a interseção se parecerá com y_a ou y_b , e nenhuma dessas duas permite um fluxo contínuo de pontas de setas. O tipo de trajetória temporal $y(t)$ correspondente a essa linha de fase em laço é ilustrado na Figura 15.4c. Note que, sempre que $y(t)$ atinge o limite superior y'_c ou o limite inferior y_c , temos $dy/dt = 0$ (extremos locais); mas esses valores certamente não representam valores de equilíbrio de y . Em termos da Figura 15.3, isso significa que nem todas as interseções entre a linha de fase e o eixo y são posições de equilíbrio.

Em suma, para o estudo da estabilidade dinâmica de equilíbrio (ou da convergência da trajetória temporal), temos a alternativa de achar a trajetória temporal em si ou, então, simplesmente inferir a trajetória por meio de sua linha de fase. Vamos ilustrar a aplicação dessa última abordagem com o modelo de crescimento de Solow. Daqui em diante, vamos denotar o valor do equilíbrio intertemporal de y por $\bar{y}$, para distingui-lo de y^* .

EXERCÍCIO 15.6

1. Construa o gráfico da linha de fase para cada uma das seguintes e discuta suas implicações qualitativas:

(a) $\frac{dy}{dt} = y - 7$

(c) $\frac{dy}{dt} = 4 - \frac{y}{2}$

(b) $\frac{dy}{dt} = 1 - 5y$

(d) $\frac{dy}{dt} = 9y - 11$

2. Construa o gráfico da linha de fase para cada uma das seguintes e interprete-a:

(a) $\frac{dy}{dt} = (y + 1)^2 - 16 \quad (y \geq 0)$

(b) $\frac{dy}{dt} = \frac{1}{2}y - y^2 \quad (y \geq 0)$

3. Dada $dy/dt = (y - 3)(y - 5) = y^2 - 8y + 15$:

- (a) Deduza que há dois níveis de equilíbrio possíveis de y , um em $y = 3$ e outro em $y = 5$.

- (b) Encontre o sinal de $\frac{d}{dy} \left(\frac{dy}{dt} \right)$ em $y = 3$ e $y = 5$, respectivamente. O que você pode inferir deles?

15.7 Modelo de crescimento de Solow

O modelo de crescimento do professor Robert Solow,^{††} agraciado com um prêmio Nobel, pretende mostrar, entre outras coisas, que a trajetória de crescimento do fio da navalha do modelo de Domar é, primordialmente, um resultado da premissa específica da função produção nele adotada e que, sob certas circunstâncias alternativas, pode não surgir a necessidade de balanceamento delicado.

[†] Isso pode surgir de uma equação diferencial de segundo grau $(dy/dt)^2 = f(y)$.

^{††} Solow, Robert M. "A Contribution to the Theory of Economic Growth". *Quarterly Journal of Economics*, fevereiro de 1956, p. 65–94.

A estrutura

No modelo de Domar, a produção é enunciada explicitamente como uma função apenas do capital: $\kappa = \rho K$ (a capacidade produtiva, ou produção potencial, é uma constante múltipla do estoque de capital). A ausência de um insumo trabalho na função produção acarreta a implicação de que o trabalho sempre está combinado com o capital segundo uma proporção *fixa*, de modo que é viável considerar explicitamente somente um desses fatores de produção. Solow, ao contrário, procura analisar o caso em que capital e trabalho podem ser combinados em proporções *variáveis*. Assim, sua função produção aparece na forma

$$Q = f(K, L) \quad (K, L > 0)$$

onde Q é produção (líquida, sem considerar depreciação), K é capital e L é trabalho – todos usados no sentido *macro*. Supõe-se que f_K e f_L são positivas (produtos marginais positivos) e que f_{KK} e f_{LL} são negativas (retornos decrescentes para cada insumo). Além do mais, a função produção f é considerada linearmente homogênea (retornos constantes de escala). Por conseqüência, é possível escrever

$$Q = Lf\left(\frac{K}{L}, 1\right) = L\phi(k) \quad \text{onde } k \equiv \frac{K}{L} \quad (15.25)$$

Em vista das hipóteses sobre os sinais de f_K e f_{KK} , a função ϕ recém-introduzida (que, note bem, tem um único argumento, k) deve ser caracterizada por uma derivada primeira positiva e uma derivada segunda negativa. Para verificar essa afirmativa, primeiro vamos lembrar, de (12.49), que

$$f_K \equiv \text{MPP}_K = \phi'(k)$$

portanto, $f_K > 0$ significa, automaticamente $\phi'(k) > 0$. Então, visto que

$$f_{KK} = \frac{\partial}{\partial K} \phi'(k) = \frac{d\phi'(k)}{dk} \frac{\partial k}{\partial K} = \phi''(k) \frac{1}{L} \quad [\text{veja (12.48)}]$$

a hipótese $f_{KK} < 0$ leva diretamente ao resultado $\phi''(k) < 0$. Assim, a função ϕ – a qual, de acordo com (12.46), resulta no APP_L para cada razão capital-trabalho – é uma função que cresce com k a uma taxa decrescente.

Dado que Q depende de K e L , agora é necessário estipular como essas duas últimas variáveis são determinadas. As premissas de Solow são:

$$\dot{K} \left(\equiv \frac{dK}{dt} \right) = sQ \quad [\text{proporção constante de } Q \text{ é investida}] \quad (15.26)$$

$$\frac{\dot{L}}{L} \left(\equiv \frac{dL/dt}{L} \right) = \lambda \quad (\lambda > 0) \quad [\text{força de trabalho cresce exponencialmente}] \quad (15.27)$$

O símbolo s representa uma propensão marginal a poupar (constante) e λ , uma taxa (constante) de crescimento do trabalho. Note a natureza dinâmica dessas premissas; elas não especificam como os *níveis* de K e L são determinados, mas como são suas *taxas de variação*.

As equações (15.25) a (15.27) constituem um modelo completo. Para resolver esse modelo, em primeiro lugar, vamos sintetizá-lo em uma única equação com uma só variável. Para começar, substitua (15.25) em (15.26) para obter

$$\dot{K} = sL\phi(k) \quad (15.28)$$

Contudo, visto que $k \equiv K/L$, e $K \equiv kL$, podemos obter uma outra expressão para $\dot{K}$ pela diferenciação da última identidade:

$$\begin{aligned}\dot{K} &= L\dot{k} + k\dot{L} && \text{[regra do produto]} \\ &= L\dot{k} + k\lambda L && \text{[por (15.27)]}\end{aligned}\quad (15.29)$$

Quando igualamos (15.29) a (15.28) e o fator comum L é eliminado, surge o resultado

$$\dot{k} = s\phi(k) - \lambda k \quad (15.30)$$

Essa equação – uma equação diferencial na variável k , com dois parâmetros s e λ – é a equação fundamental do modelo de crescimento de Solow.

Uma análise gráfico-qualitativa

Como (15.30) é enunciada sob a forma de uma função geral, não há nenhuma solução quantitativa disponível. Não obstante, podemos analisá-la qualitativamente. Para essa finalidade, devemos representar graficamente uma linha de fase, com $\dot{k}$ no eixo vertical e k no horizontal.

Uma vez que (15.30) contém dois termos na direita, contudo, primeiramente vamos desenhar os gráficos das duas curvas em separado. É óbvio que o termo λk , uma função linear de k , aparecerá na Figura 15.5a como uma linha reta, com interseção vertical no zero e uma inclinação igual a λ . A representação gráfica do termo $s\phi(k)$, por outro lado, é uma curva que cresce a uma taxa decrescente, como $\phi(k)$, já que $s\phi(k)$ é uma mera fração constante da curva $\phi(k)$. Se considerarmos K como um fator de produção indispensável, devemos iniciar a curva $s\phi(k)$ a partir do ponto de origem; isso porque, se $K = 0$ e, assim, $k = 0$, Q também deve ser zero, bem como o serão $\phi(k)$ e $s\phi(k)$. Na verdade, o modo como a curva se apresenta também reflete a premissa implícita de que existe um conjunto de valores k para os quais $s\phi(k)$ excede λk , de modo que as duas curvas se interceptam em algum valor positivo de k , a saber, $\bar{k}$.

Com base nessas curvas, o valor de $\dot{k}$ para cada valor de k pode ser medido pela distância vertical entre as duas curvas. Desenhar os valores de $\dot{k}$ em função de k , como na Figura 15.5b, nos dará, então, a linha de fase de que precisamos. Note que, uma vez que as duas curvas na Figura 15.5a se interceptam quando a razão capital-trabalho for $\bar{k}$, a linha de fase na Figura 15.5b deve cruzar o eixo horizontal em $\bar{k}$. Isso marca $\bar{k}$ como a razão capital-trabalho de equilíbrio intertemporal.

Considerando que a linha de fase tem uma inclinação negativa em $\bar{k}$, o equilíbrio é imediatamente identificado como estável; dado qualquer valor inicial (positivo) de k , o movimento dinâmico

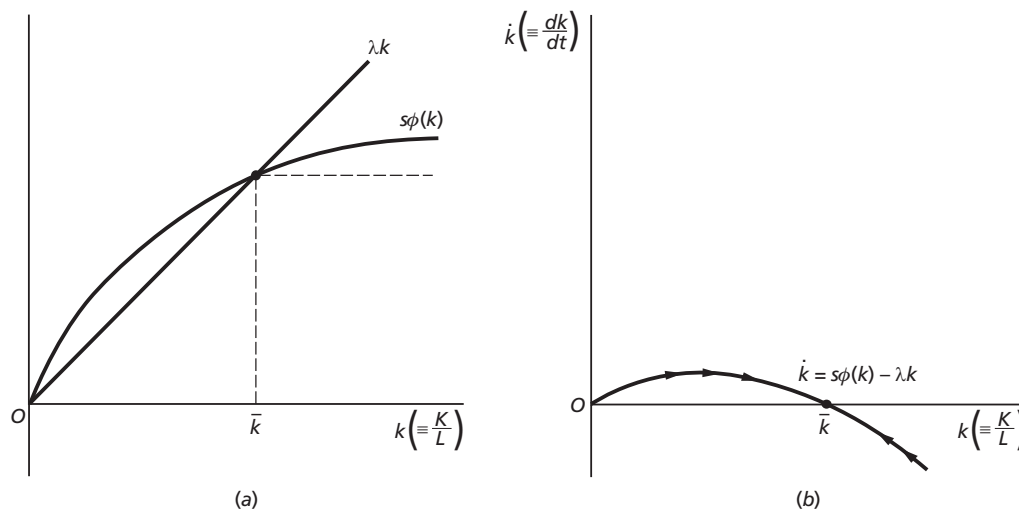


FIGURA 15.5

mico do modelo deve nos levar, por convergência, ao nível de equilíbrio $\bar{k}$. O ponto significativo é que, uma vez alcançado esse equilíbrio – e, por isso a razão capital-trabalho (por definição) é invariável ao longo do tempo –, dali em diante o capital deve crescer par a par com o trabalho, a uma taxa idêntica λ . Isso implicará, por sua vez, que o investimento líquido tem de crescer à taxa λ (veja Exercício 15.7–2). Note, contudo, que a palavra *tem* é usada aqui não no sentido de requisito, mas no sentido de automatismo. Assim, o modelo de Solow serve para mostrar que, dada uma taxa de crescimento do trabalho λ , a economia, por si só, e sem o delicado balanceamento ao modo de Domar, pode, em seu devido tempo, chegar a um estado de crescimento constante no qual o investimento crescerá à taxa λ , a mesma de K e L . Além do mais, para satisfazer (15.25), Q também deve crescer à mesma taxa, porque $\phi(k)$ é uma constante quando a razão capital-trabalho permanecer invariável no nível $\bar{k}$. Tal situação, na qual todas as variáveis relevantes crescem a uma taxa idêntica, é denominada um *estado estacionário* – uma generalização do conceito de *estado constante* (no qual todas as variáveis relevantes permanecem constantes ou, em outras palavras, todas crescem à taxa zero).

Note que, na análise precedente, admite-se, por conveniência, que a função produção é invariável ao longo do tempo. Por outro lado, se for permitido melhorar o estado da tecnologia, a função produção terá de ser devidamente modificada. Por exemplo, ela pode ser escrita, alternativamente, na forma

$$Q = T(t) f(K, L) \quad \left(\frac{dT}{dt} > 0 \right)$$

onde T , algum tipo de medida da tecnologia, é uma função crescente do tempo. Por causa do termo multiplicativo crescente $T(t)$, uma quantidade fixa de K e L renderá, em data futura, uma produção maior que no presente. Nesse caso, a curva $\phi(k)$ na Figura 15.5 estará sujeita a um movimento secular (temporal) para cima, resultando em interseções sucessivamente mais altas com o raio λk e também em valores maiores de k . Por conseguinte, com a melhoria tecnológica e por uma sucessão de estados estacionários, será possível ter uma quantidade cada vez maior de equipamento primordial à disposição de cada trabalhador representativo na economia, com uma elevação concomitante da produtividade.

Uma ilustração quantitativa

A análise precedente teve de ser qualitativa devido à presença de uma função geral $\phi(k)$ no modelo. Mas, se especificarmos que a função produção é uma função de Cobb-Douglas linearmente homogênea, por exemplo, então também podemos achar uma solução quantitativa.

Vamos escrever a função produção como

$$Q = K^\alpha L^{1-\alpha} = L \left(\frac{K}{L} \right)^\alpha = L k^\alpha$$

de modo que $\phi(k) = k^\alpha$. Então, (15.30) torna-se

$$\dot{k} = s k^\alpha - \lambda k \quad \text{ou} \quad \dot{k} + \lambda k = s k^\alpha$$

que é uma equação de Bernoulli na variável k [veja (15.24)], com $R = \lambda$, $T = s$, e $m = \alpha$. Fazendo $z = k^{1-\alpha}$, obtemos sua versão linearizada

$$dz + [(1-\alpha)\lambda z - (1-\alpha)s] dt = 0$$

ou

$$\frac{dz}{dt} + \underbrace{(1-\alpha)\lambda}_{a} z = \underbrace{(1-\alpha)s}_{b}$$

Essa é uma equação diferencial linear com um coeficiente constante a e um termo constante b . Assim, pela fórmula (15.5'), temos

$$z(t) = \left[z(0) - \frac{s}{\lambda} \right] e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

A substituição de $z = k^{1-\alpha}$ então resultará na solução final

$$k^{1-\alpha} = \left[k(0)^{1-\alpha} - \frac{s}{\lambda} \right] e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

onde $k(0)$ é o valor inicial da razão capital-trabalho k .

Essa solução é o que determina a trajetória temporal de k . Lembrando que $(1 - \alpha)$ e λ são ambas positivas, vemos que, quando $t \rightarrow \infty$, a expressão exponencial se aproxima de zero; conseqüentemente,

$$k^{1-\alpha} \rightarrow \frac{s}{\lambda} \quad \text{ou} \quad k \rightarrow \left(\frac{s}{\lambda} \right)^{1/(1-\alpha)} \quad \text{quando } t \rightarrow \infty$$

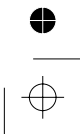
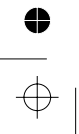
Portanto, a razão capital-trabalho se aproximará de uma constante como seu valor de equilíbrio. Esse equilíbrio, ou valor de estado estacionário, $(s/\lambda)^{1/(1-\alpha)}$, varia diretamente com a propensão a poupar s e inversamente com a taxa de crescimento do trabalho λ .

EXERCÍCIO 15.7

1. Divida todos os termos de (15.30) por k e interprete a equação resultante em termos das taxas de crescimento de k , K e L .
2. Mostre que, se o capital estiver crescendo à taxa λ (isto é, $K = Ae^{\lambda t}$), o investimento líquido I também deve estar crescendo àquela taxa λ .
3. As variáveis insumos originais do modelo de Solow são K e L , mas, em vez disso, a equação fundamental (15.30) focaliza a razão capital-trabalho k . Qual premissa (ou premissas) no modelo é responsável por essa mudança de foco e a torna possível? Explique.
4. Desenhe um diagrama de fase para cada uma das seguintes e discuta os aspectos qualitativos da trajetória temporal $y(t)$:

(a) $\dot{y} = 3 - y - \ln y$

(b) $\dot{y} = e^y - (y + 2)$



CAPÍTULO 16

Equações diferenciais de ordem mais alta

No Capítulo 15, discutimos os métodos para resolver uma equação diferencial de *primeira ordem*, na qual não aparece nenhuma derivada (ou diferencial) de ordens mais altas que 1. Entretanto, às vezes a especificação de um modelo pode envolver a derivada segunda ou uma derivada de uma ordem ainda mais alta. Podemos, por exemplo, encontrar uma função que descreve “a taxa de variação da taxa de variação” da variável de renda Y , digamos,

$$\frac{d^2 Y}{dt^2} = kY$$

pela qual teríamos de achar a trajetória temporal de Y . Nesse caso, a função dada constitui uma equação diferencial de *segunda ordem* e a tarefa de achar a trajetória temporal $Y(t)$ equivale a *resolver* a equação diferencial de segunda ordem. O presente capítulo aborda os métodos de solução e as aplicações econômicas dessas equações diferenciais de ordem mais alta, mas restringiremos nossa discussão apenas ao caso das equações *lineares*.

Uma variedade simples de equações diferenciais lineares de ordem n tem a seguinte forma:

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_{n-1} \frac{dy}{dt} + a_n y = b \quad (16.1)$$

ou, adotando uma notação alternativa,

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y'(t) + a_n y = b \quad (16.1')$$

Essa equação é de ordem n porque a n -ésima derivada (o primeiro termo da esquerda) é a derivada mais alta presente. É *linear*, já que todas as derivadas, bem como a variável dependente, aparecem somente no primeiro grau e, além do mais, não ocorre nenhum termo de produto no qual y é multiplicado por qualquer de suas derivadas. Além disso, você notará que essa equação diferencial é caracterizada por *coeficientes constantes* (os a 's) e um *termo constante* (b). A constância dos coeficientes é uma premissa que adotaremos em todo este capítulo. O termo constante b , por outro lado, é adotado aqui como uma primeira aproximação; mais adiante, na Seção 16.5, nós o descartaremos em favor de um termo variável.

16.1 Equações diferenciais lineares de segunda ordem com coeficientes constantes e termo constante

Por questões pedagógicas, vamos discutir, em primeiro lugar, o método de solução para o caso de *segunda ordem* ($n = 2$). Então, a equação diferencial relevante é a equação simples:

$$y''(t) + a_1 y'(t) + a_2 y = b \quad (16.2)$$

onde a_1 , a_2 e b são todos constantes. Se o termo b for identicamente zero, temos uma equação *homogênea*, mas se b for uma constante diferente de zero, a equação é *não-homogênea*. Nossa discussão continuará tendo como premissa que (16.2) é não-homogênea; no decurso da solução da versão não-homogênea de (16.2) surgirá automaticamente a solução da versão homogênea como um subproduto.

Neste contexto, lembramos uma proposição apresentada na Seção 15.1 que é igualmente aplicável aqui: se y_c for a *função complementar*, isto é, a solução geral (contendo constantes arbitrárias) da equação homogênea associada a (16.2) e se y_p for a *solução particular*, isto é, qualquer solução particular (que não contém qualquer constante arbitrária) da equação completa (16.2), então $y(t) = y_c + y_p$ será a solução geral da equação completa. Como já explicado anteriormente, o componente y_p nos dá o valor de equilíbrio da variável y no sentido intertemporal do termo, ao passo que o componente y_c revela, em cada instante, o desvio da trajetória temporal $y(t)$ em relação ao equilíbrio.

A solução particular

Para o caso de coeficientes constantes e termo constante, a solução particular é relativamente fácil de achar. Visto que a solução particular pode ser *qualquer* solução de (16.2), isto é, qualquer valor de y que satisfaça essa equação não-homogênea, devemos sempre tentar o tipo mais simples possível, a saber, $y = \text{uma constante}$. Se $y = \text{uma constante}$, segue que

$$y'(t) = y''(t) = 0$$

de modo que (16.2) torna-se, na verdade, $a_2 y = b$, com a solução $y = b/a_2$. Assim, a solução particular desejada é

$$y_p = \frac{b}{a_2} \quad (\text{caso de } a_2 \neq 0) \quad (16.3)$$

Uma vez que o processo de encontrar o valor de y_p envolve a condição $y'(t) = 0$, o princípio racional para considerar aquele valor como um equilíbrio intertemporal fica evidente por si só.

EXEMPLO 1 Encontre uma solução particular da equação

$$y''(t) + y'(t) - 2y = -10$$

Aqui, os coeficientes relevantes são $a_2 = -2$ e $b = -10$. Portanto, a solução particular é $y_p = -10/(-2) = 5$.

E se $a_2 = 0$ – de modo que a expressão b/a_2 não está definida? Nessa situação, uma vez que a solução constante para y_p não funciona, temos de experimentar alguma forma *não-constante* de solução. Optando pela possibilidade mais simples, podemos tentar $y = kt$. Visto que $a_2 = 0$, a equação diferencial agora é

$$y''(t) + a_1 y'(t) = b$$



mas, se $y = kt$, o que implica $y'(t) = k$ e $y''(t) = 0$, essa equação se reduz a $a_1 k = b$. Isso determina o valor de k como b/a_1 , o que nos dá, desse modo, a solução particular

$$y_p = \frac{b}{a_1} t \quad (\text{caso de } a_2 = 0; a_1 \neq 0) \quad (16.3')$$

Considerando que y_p , nesse caso, é uma função não-constante do tempo, vamos considerá-la como um equilíbrio móvel.

EXEMPLO 2

Encontre y_p para a equação $y''(t) + y'(t) = -10$. Aqui, temos $a_2 = 0$, $a_1 = 1$ e $b = -10$. Assim, por (16.3'), podemos escrever

$$y_p = -10t$$

Se acontecer de a_1 também ser zero, então a forma de solução de $y = kt$ também falhará porque a expressão bt/a_1 agora estará indefinida. Então, devemos tentar uma solução da forma $y = kt^2$. Com $a_1 = a_2 = 0$, agora a equação diferencial se reduz à forma extremamente simples

$$y''(t) = b$$

e, se $y = kt^2$, o que implica $y'(t) = 2kt$ e $y''(t) = 2k$, a equação diferencial pode ser escrita como $2k = b$. Assim, achamos $k = b/2$, e a solução particular é

$$y_p = \frac{b}{2} t^2 \quad (\text{caso de } a_1 = a_2 = 0) \quad (16.3'')$$

O equilíbrio representado por essa integral particular é novamente um equilíbrio móvel.

EXEMPLO 3

Encontre y_p para a equação $y''(t) = -10$. Visto que os coeficientes são $a_1 = a_2 = 0$ e $b = -10$, a fórmula (16.3'') é aplicável. A resposta desejada é $y_p = -5t^2$.

A função complementar

A função complementar de (16.2) é definida como a solução geral de sua equação associada homogênea

$$y''(t) + a_1 y'(t) + a_2 y = 0 \quad (16.4)$$

É por isso que afirmamos que a solução de uma equação homogênea sempre será um *subproduto* do processo de resolução de uma equação completa.

Ainda que nunca tenhamos enfrentado tal equação antes, nossa experiência com a função complementar das equações diferenciais de primeira ordem pode nos dar uma indicação útil. Pelas soluções (15.3), (15.3'), (15.5) e (15.5'), fica claro que as expressões exponenciais da forma Ae^{rt} têm um lugar muito proeminente nas funções complementares de equações diferenciais de primeira ordem com coeficientes constantes. Então, por que não tentar uma solução da forma $y = Ae^{rt}$ na equação de segunda ordem também?

Se adotarmos a solução experimental $y = Ae^{rt}$, também devemos aceitar

$$y'(t) = rAe^{rt} \quad \text{e} \quad y''(t) = r^2 Ae^{rt}$$

como as derivadas de y . Com base nessas expressões para y , $y'(t)$ e $y''(t)$, a equação diferencial homogênea (16.4) pode ser transformada em

$$Ae^{rt} (r^2 + a_1 r + a_2) = 0 \quad (16.4')$$

Contanto que escolhamos os valores de A e r que satisfazem (16.4'), a solução experimental $y = Ae^{rt}$ deveria funcionar. Uma vez que e^{rt} nunca pode ser zero, devemos fazer ou $A = 0$ ou providenciar que r satisfaça a equação

$$r^2 + a_1 r + a_2 = 0 \quad (16.4'')$$

Uma vez que o valor da constante (arbitrária) A deve ser definido por meio da utilização das condições iniciais do problema, contudo, não podemos simplesmente estabelecer $A = 0$ à vontade. Por conseguinte, é essencial procurar valores de r que satisfaçam (16.4'').

A equação (16.4'') é conhecida como a *equação característica* (ou *equação auxiliar*) da equação homogênea (16.4), ou da equação completa (16.2). Por ser uma equação quadrática em r , ela dá duas raízes (soluções) às quais nos referimos no presente contexto como *raízes características*, como se segue:[†]

$$r_1, r_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \quad (16.5)$$

Essas duas raízes guardam entre si uma relação simples, mas interessante, que pode servir como um meio conveniente de verificar nosso cálculo: a *soma* das duas raízes é sempre igual a $-a_1$ e seu *produto* é sempre igual a a_2 . A demonstração dessa afirmação é direta:

$$\begin{aligned} r_1 + r_2 &= \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2} + \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2} = \frac{-2a_1}{2} = -a_1 \\ r_1 r_2 &= \frac{(-a_1)^2 - (a_1^2 - 4a_2)}{4} = \frac{4a_2}{4} = a_2 \end{aligned} \quad (16.6)$$

Os valores dessas duas raízes são os únicos valores que podemos atribuir a r na solução $y = Ae^{rt}$. Mas isso significa que, na verdade, há *duas* soluções que funcionarão, a saber,

$$y_1 = A_1 e^{r_1 t} \quad \text{e} \quad y_2 = A_2 e^{r_2 t}$$

onde A_1 e A_2 são duas constantes arbitrárias e r_1 e r_2 são as raízes características dadas por (16.5). Uma vez que queremos apenas *uma* solução geral, entretanto, parece que uma delas está sobrando. Agora há duas alternativas à nossa disposição: (1) escolher y_1 ou y_2 aleatoriamente; ou (2) combiná-las de algum modo.

A primeira alternativa, embora mais simples, é inaceitável. Há somente uma constante arbitrária em y_1 ou y_2 , mas, para se qualificar como uma solução geral de uma equação diferencial de *segunda ordem*, a expressão deve conter *duas* constantes arbitrárias. Esse requisito origina-se do fato de que, ao passarmos de uma função $y(t)$ para sua derivada segunda $y''(t)$, “perdemos” duas constantes durante as duas rodadas de diferenciação; portanto, para voltar de uma equação diferencial de segunda ordem para a função primitiva $y(t)$, devemos reincorporar duas constantes. Isso nos deixa apenas a alternativa de combinar y_1 e y_2 de modo a incluir ambas as constantes A_1 e A_2 . No fim, podemos simplesmente tomar sua *soma*, $y_1 + y_2$, como a solução geral de (16.4). Vamos demonstrar que, se y_1 e y_2 , respectivamente, satisfizerem (16.4), então a soma ($y_1 + y_2$) também o fará. Se y_1 e y_2 forem, de fato, soluções de (16.4), então, substituindo cada uma delas em (16.4), devemos constatar que as duas equações seguintes valem:

$$y_1''(t) + a_1 y_1'(t) + a_2 y_1 = 0$$

$$y_2''(t) + a_1 y_2'(t) + a_2 y_2 = 0$$

[†] Note que a equação quadrática (16.4'') está na forma normalizada; o coeficiente do termo r^2 é 1. Ao aplicar a fórmula (16.5) para achar as raízes características de uma equação diferencial, primeiro devemos nos certificar de que a equação característica está, de fato, na forma normalizada.

Contudo, somando essas duas equações, constatamos que

$$\underbrace{[y_1''(t)] + y_2''(t)}_{=\frac{d^2}{dt^2}(y_1 + y_2)} + a_1 \underbrace{[y_1'(t)] + y_2'(t)}_{=\frac{d}{dt}(y_1 + y_2)} + a_2(y_1 + y_2) = 0$$

Portanto, assim como y_1 ou y_2 , a soma $(y_1 + y_2)$ também satisfaz a equação (16.4). De acordo com isso, a solução geral da equação homogênea (16.4) ou a função complementar da equação completa (16.2) pode ser escrita, em geral, como $y_c = y_1 + y_2$.

Entretanto, um exame mais cuidadoso da fórmula da raiz característica (16.5) indica que, no que diz respeito aos valores de r_1 e r_2 , podem surgir três casos possíveis, alguns dos quais podem exigir uma modificação de nosso resultado $y_c = y_1 + y_2$.

Caso 1 (raízes reais distintas) Quando $a_1^2 > 4a_2$, a raiz quadrada em (16.5) é um número real, e as duas raízes r_1 e r_2 assumirão valores reais *distintos*, porque a raiz quadrada é adicionada a $-a_1$ para r_1 , mas subtraída de $-a_1$ para r_2 . Nesse caso, podemos, de fato, escrever

$$y_c = y_1 + y_2 = A_1 e^{r_1 t} + A_2 e^{r_2 t} \quad (r_1 \neq r_2) \quad (16.7)$$

Como as duas raízes são distintas, as duas expressões exponenciais devem ser linearmente independentes (nenhuma é múltipla da outra); por consequência, A_1 e A_2 permanecerão sempre como entidades separadas e nos darão duas constantes, como exigido.

EXEMPLO 4 Resolva a equação diferencial

$$y''(t) + y'(t) - 2y = -10$$

No Exemplo 1, já foi constatado que a solução particular dessa equação é $y_p = 5$. Vamos encontrar a função complementar. Uma vez que os coeficientes da equação são $a_1 = 1$ e $a_2 = -2$, as raízes características são, por (16.5),

$$r_1, r_2 = \frac{-1 \pm \sqrt{1+8}}{2} = \frac{-1 \pm 3}{2} = 1, -2$$

(Verifique: $r_1 + r_2 = -1 = -a_1$; $r_1 r_2 = -2 = a_2$) Visto que as raízes são números reais distintos, a função complementar é $y_c = A_1 e^t + A_2 e^{-2t}$. Por conseguinte, a solução geral pode ser escrita como

$$y(t) = y_c + y_p = A_1 e^t + A_2 e^{-2t} + 5 \quad (16.8)$$

Para definir as constantes A_1 e A_2 , agora são necessárias *duas* condições iniciais. Sejam essas duas condições $y(0) = 12$ e $y'(0) = -2$. Isto é, quando $t = 0$, $y(t)$ e $y'(t)$ são, respectivamente, 12 e -2 . Fazendo $t = 0$ em (16.8), constatamos que

$$y(0) = A_1 + A_2 + 5$$

Diferenciando (16.8) em relação a t e estabelecendo $t = 0$ na derivada, constatamos que

$$y'(t) = A_1 e^t - 2A_2 e^{-2t} \quad \text{e} \quad y'(0) = A_1 - 2A_2$$

Por conseguinte, para satisfazer as duas condições iniciais, devemos estabelecer $y(0) = 12$ e $y'(0) = -2$, o que resulta no seguinte par de equações simultâneas:

$$A_1 + A_2 = 7$$

$$A_1 - 2A_2 = -2$$

com soluções $A_1 = 4$ e $A_2 = 3$. Assim, a solução definida da equação diferencial é

$$y(t) = 4e^t + 3e^{-2t} + 5 \quad (16.8')$$

Como antes, podemos verificar a validade dessa solução por diferenciação. As derivadas primeira e segunda de (16.8') são

$$y'(t) = 4e^t - 6e^{-2t} \quad \text{e} \quad y''(t) = 4e^t + 12e^{-2t}$$

Quando essas duas expressões são substituídas na equação diferencial dada, juntamente com (16.8'), o resultado é uma identidade $-10 = -10$. Assim, a solução está correta. Como você pode verificar com facilidade, (16.8') também satisfaz ambas as condições iniciais.

Caso 2 (raízes reais repetidas) Quando os coeficientes da equação diferencial são tais que $a_1^2 = 4a_2$, a raiz quadrada em (16.5) desaparecerá e as duas raízes características assumirão um valor idêntico:

$$r (= r_1 = r_2) = -\frac{a_1}{2}$$

Essas raízes são conhecidas como *raízes repetidas* ou *múltiplas* (aqui, *duplas*).

Se tentarmos escrever a função complementar como $y_c = y_1 + y_2$, nesse caso a soma será reduzida para uma única expressão

$$y_c = A_1 e^{rt} + A_2 e^{rt} = (A_1 + A_2) e^{rt} = A_3 e^{rt}$$

o que nos deixa com somente uma constante. Isso não é suficiente para nos reconduzir de uma equação diferencial de segunda ordem até sua função primitiva. A única saída é achar um outro termo componente candidato à soma – um termo que satisfaça (16.4) e que, ainda assim, seja linearmente independente do termo $A_3 e^{rt}$, de modo a impedir essa “redução”.

Uma expressão que cumprirá esses requisitos é $A_4 t e^{rt}$. Visto que a variável t entra nessa expressão sob a forma multiplicativa, esse termo componente é obviamente linearmente independente do termo $A_3 e^{rt}$; assim, ele nos habilitará a introduzir uma outra constante, A_4 . Mas $A_4 t e^{rt}$ se qualifica como uma solução de (16.4)? Se tentarmos $y = A_4 t e^{rt}$, então, pela regra do produto, podemos constatar que suas derivadas primeira e segunda são

$$y'(t) = (rt + 1)A_4 e^{rt} \quad \text{e} \quad y''(t) = (r^2 t + 2r)A_4 e^{rt}$$

Substituindo essas expressões de y , y' e y'' no lado esquerdo de (16.4), obtemos a expressão

$$[(r^2 t + 2r) + a_1(rt + 1) + a_2 t]A_4 e^{rt}$$

Considerando que, no presente contexto, temos $a_1^2 = 4a_2$ e $r = -a_1/2$, esta última expressão desaparece identicamente e, assim, é sempre igual ao lado direito de (16.4); isso mostra que $A_4 t e^{rt}$ realmente se qualifica como uma solução.

Portanto, a função complementar do caso de raízes duplas pode ser escrita como

$$y_c = A_3 e^{rt} + A_4 t e^{rt} \quad (16.9)$$

EXEMPLO 5 Resolva a equação diferencial

$$y''(t) + 6y'(t) + 9y = 27$$

Aqui, os coeficientes são $a_1 = 6$ e $a_2 = 9$; já que $a_1^2 = 4a_2$, as raízes serão repetidas. Segundo a fórmula (16.5), temos $r = -a_1/2 = -3$. Assim, de acordo com o resultado em (16.9), a função complementar pode ser escrita como

$$y_c = A_3 e^{-3t} + A_4 t e^{-3t}$$

A solução geral da equação diferencial dada agora pode ser obtida de imediato. Experimentando uma solução constante para a solução particular, obtemos $y_p = 3$. Segue-se que a solução geral da equação completa é

$$y(t) = y_c + y_p = A_3 e^{-3t} + A_4 t e^{-3t} + 3$$

Mais uma vez, as duas constantes arbitrárias podem ser definidas com duas condições iniciais. Suponha que as condições iniciais são $y(0) = 5$ e $y'(0) = -5$. Fazendo $t = 0$ na solução geral precedente, devemos achar $y(0) = 5$; isto é,

$$y(0) = A_3 + 3 = 5$$

Isso dá como resultado $A_3 = 2$. Em seguida, diferenciando a solução geral e fazendo $t = 0$ e também $A_3 = 2$, devemos ter $y'(0) = -5$. Isto é,

$$y'(t) = -3A_3 e^{-3t} - 3A_4 t e^{-3t} + A_4 e^{-3t}$$

$$\text{e} \quad y'(0) = -6 + A_4 = -5$$

Isso dá como resultado $A_4 = 1$. Assim, podemos finalmente escrever a solução definida da equação dada como

$$y(t) = 2e^{-3t} + te^{-3t} + 3$$

Caso 3 (raízes complexas) Resta uma terceira possibilidade em relação à grandeza relativa dos coeficientes a_1 e a_2 , a saber, $a_1^2 < 4a_2$. Quando ocorrer essa eventualidade, a fórmula (16.5) envolverá a raiz quadrada de um número *negativo*, que não pode ser manuseada antes de apresentarmos adequadamente os conceitos de números *imaginários* e *complexos*. Por enquanto, portanto, nos contentaremos com a mera citação desse caso e deixaremos para discuti-lo com mais profundidade nas Seções 16.2 e 16.3.

Os três casos citados podem ser ilustrados pelas três curvas na Figura 16.1, cada uma delas representando uma versão diferente da forma quadrática $f(r) = r^2 + a_1 r + a_2$. Como já aprendemos antes, quando tal função é igualada a zero, o resultado é a *equação quadrática* $f(r) = 0$ e resolver essa equação é simplesmente “achar os zeros da *função quadrática*”. Em termo gráficos, isso significa que as raízes da equação serão encontradas no eixo horizontal, onde $f(r) = 0$.

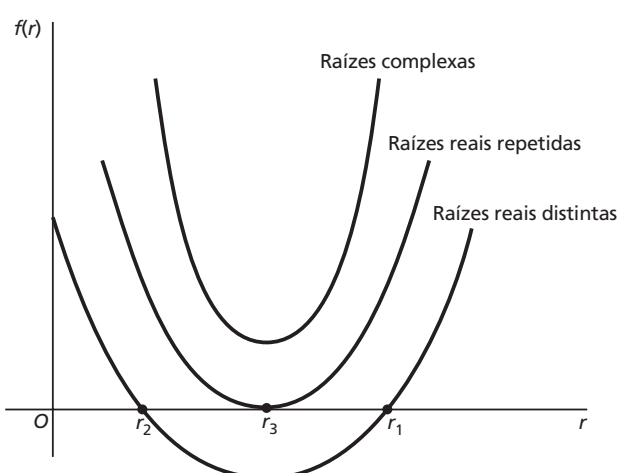
A posição da curva que está mais embaixo na Figura 16.1 é tal que ela intercepta o eixo horizontal duas vezes; assim, podemos achar duas raízes distintas r_1 e r_2 , ambas as quais satisfazem a equação quadrática $f(r) = 0$ e ambas as quais, é claro, têm valor real. Assim, a curva que está mais embaixo ilustra o Caso 1. Considerando a curva do meio, observamos que ela toca o eixo horizontal apenas uma vez, em r_3 . Esse é o único valor de r que pode satisfazer a equação $f(r) = 0$. Portanto, a curva do meio ilustra o Caso 2. Por último, notamos que a curva mais acima não encontra o eixo horizontal de jeito nenhum e, por isso, não há nenhuma raiz de valor real para a equação $f(r) = 0$. Conquanto não exista nenhuma raiz real neste caso, ainda assim há dois números complexos que podem satisfazer a equação, como será mostrado na Seção 16.2.

A estabilidade dinâmica de equilíbrio

Para os Casos 1 e 2, a condição para estabilidade dinâmica de equilíbrio depende novamente dos sinais algébricos das raízes características.

Para o Caso 1, a função complementar (16.7) consiste nas duas expressões exponenciais $A_1 e^{r_1 t}$ e $A_2 e^{r_2 t}$. Os coeficientes A_1 e A_2 são constantes arbitrárias; seus valores dependem das condições iniciais do problema. Assim, podemos ter certeza de um equilíbrio dinamicamente estável ($y_c \rightarrow 0$ quando $t \rightarrow \infty$), independentemente de quais sejam as condições iniciais, se, e somente

FIGURA 16.1



se, as raízes r_1 e r_2 forem *ambas* negativas. Enfatizamos a palavra *ambas* aqui, porque a condição para estabilidade dinâmica *não* permite que nem mesmo *uma* das raízes seja positiva ou zero. Se $r_1 = 2$ e $r_2 = -5$, por exemplo, poderá parecer, à primeira vista, que a segunda raiz, por ser maior em valor absoluto, pode ser mais importante que a primeira. Contudo, na verdade, é a raiz *positiva* que deve eventualmente dominar porque, à medida que t aumenta, e^{2t} fica cada vez maior, mas e^{-5t} míngua a um passo constante.

Para o Caso 2, com raízes repetidas, a função complementar (16.9) contém não somente a expressão familiar e^{rt} , mas também uma expressão multiplicativa te^{rt} . Para que o primeiro termo se aproxime de zero sejam quais forem as condições iniciais, é necessário e suficiente que $r < 0$. Mas isso também asseguraria o desaparecimento de te^{rt} ? No fim, a expressão te^{rt} (ou, em termos mais gerais, $t^k e^{rt}$) possui o mesmo tipo geral de trajetória temporal que e^{rt} ($r \neq 0$). Assim, a condição $r < 0$ é, de fato, necessária e suficiente para que toda a função complementar se aproxime de zero quando $t \rightarrow \infty$, resultando em um equilíbrio intertemporal dinamicamente estável.

EXERCÍCIO 16.1

- Encontre a solução particular de cada equação:

(a) $y''(t) - 2y'(t) + 5y = 2$	(d) $y''(t) + 2y'(t) - y = -4$
(b) $y''(t) + y'(t) = 7$	(e) $y''(t) = 12$
(c) $y''(t) + 3y = 9$	
- Encontre a função complementar de cada equação:

(a) $y''(t) + 3y'(t) - 4y = 12$	(c) $y''(t) - 2y'(t) + y = 3$
(b) $y''(t) + 6y'(t) + 5y = 10$	(d) $y''(t) + 8y'(t) + 16y = 0$
- Encontre a solução geral de cada equação diferencial no Problema 3, e então defina a solução com as condições iniciais $y(0) = 4$ e $y'(0) = 2$.
- Os equilíbrios intertemporais encontrados no Problema 2 são dinamicamente estáveis?
- Verifique se a solução definida no Exemplo 5 de fato (a) satisfaz as duas condições iniciais e (b) tem derivadas primeira e segunda que satisfazem a equação diferencial dada.
- Mostre que, quando $t \rightarrow \infty$, o limite de te^{rt} é zero se $r < 0$, mas é infinito se $r \geq 0$.

16.2 Números complexos e funções circulares

Quando os coeficientes de uma equação diferencial linear de segunda ordem $y''(t) + a_1 y'(t) + a_2 y = b$ são tais que $a_1^2 < 4a_2$, a fórmula da raiz característica (16.5) exigiria o cálculo da raiz quadrada de um número *negativo*. Visto que o quadrado de qualquer número real positivo ou negativo é invariavelmente positivo, enquanto o quadrado de zero é zero, somente um número real *não-negativo* poderia jamais dar como resultado uma raiz quadrada de valor real. Assim, se restringirmos nossa atenção ao sistema de números reais, como fizemos até agora, não há nenhuma raiz característica disponível para esse caso (Caso 3). Esse fato nos motiva a considerar números fora do sistema de números reais.

Números imaginários e complexos

Em termos conceituais, é possível definir um número $i \equiv \sqrt{-1}$ que, quando elevado ao quadrado, será igual a -1 . Como i é a raiz quadrada de um número negativo, é óbvio que ele não tem valor real; portanto, é denominado um *número imaginário*. Com esse número à nossa disposição, podemos escrever uma profusão de outros números imaginários, tais como $\sqrt{-9} = \sqrt{9}\sqrt{-1} = 3i$ e $\sqrt{-2} = \sqrt{2}i$.

Ampliando sua aplicação um pouco mais, podemos construir ainda um outro tipo de número – um tipo que contém uma parte *real*, bem como uma parte *imaginária*, tal como $(8 + i)$ e $(3 + 5i)$. Conhecidos como *números complexos*, eles podem ser representados de modo mais geral na forma $(h + vi)$, onde h e v são dois números reais.[†] É claro que, caso $v = 0$, o número complexo será reduzido a um número real, enquanto, se $h = 0$, ele se tornará um número imaginário. Assim, o conjunto de todos os números reais (denominado $\mathbf{R}$) constitui um subconjunto do conjunto de todos os números complexos (denominado $\mathbf{C}$). De modo semelhante, o conjunto de todos os números imaginários (denominado $\mathbf{I}$) também constitui um subconjunto de $\mathbf{C}$. Isto é, $\mathbf{R} \subset \mathbf{C}$ e $\mathbf{I} \subset \mathbf{C}$. Além disso, visto que os termos *real* e *imaginário* são mutuamente exclusivos, os conjuntos $\mathbf{R}$ e $\mathbf{I}$ devem ser disjuntos, isto é, $\mathbf{R} \cap \mathbf{I} = \emptyset$.

Um número complexo $(h + vi)$ pode ser representado graficamente no que denominamos um *diagrama de Argand*, como ilustrado na Figura 16.2. Colocando h horizontalmente no eixo *real* e v verticalmente no eixo *imaginário*, o número $(h + vi)$ pode ser especificado pelo ponto (h, v) , que rotulamos alternativamente de C . Os valores de h e v têm sinais algébricos, é claro, de modo que, se $h < 0$, o ponto C estará à esquerda do ponto de origem; de modo semelhante, um v negativo significará uma localização abaixo do eixo horizontal.

Dados os valores de h e v , também podemos calcular o comprimento da linha OC aplicando o teorema de Pitágoras, que afirma que o quadrado da hipotenusa de um triângulo retângulo é a soma dos quadrados dos dois outros lados. Denotando o comprimento de OC por R (para vetor raio), temos

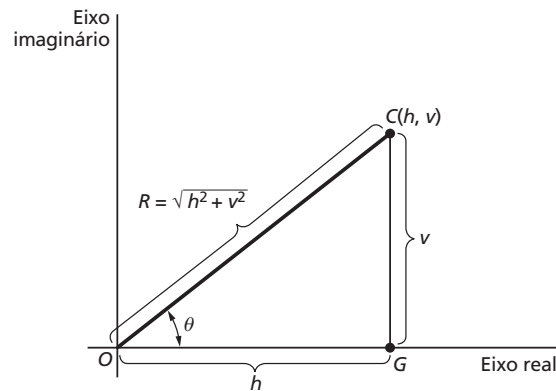


FIGURA 16.2

[†] Empregamos os símbolos h (para horizontal) e v (para vertical) na notação geral de números complexos porque logo adiante vamos desenhar os valores de h e v , respectivamente, nos eixos vertical e horizontal de um diagrama bidimensional.

$$R^2 = h^2 + v^2 \quad \text{e} \quad R = \sqrt{h^2 + v^2} \quad (16.10)$$

onde tomamos a raiz quadrada sempre como positiva. O valor de R às vezes é denominado *valor absoluto* ou *módulo* do número complexo $(h + vi)$. (Note que mudar os sinais de h e v não produzirá nenhum efeito sobre o valor absoluto do número complexo, R .) Então, assim como h e v , R tem valor real, mas, diferente desses outros valores, R é sempre positivo. Veremos que o número R é de grande importância na discussão a seguir.

Raízes complexas

Por enquanto, vamos voltar à fórmula (16.5) e examinar o caso de raízes características complexas. Quando os coeficientes de uma equação diferencial de segunda ordem são tais que $a_1^2 < 4a_2$, a raiz quadrada da expressão em (16.5) pode ser escrita como

$$\sqrt{a_1^2 - 4a_2} = \sqrt{4a_2 - a_1^2} \sqrt{-1} = \sqrt{4a_2 - a_1^2} i$$

Então, se adotarmos a notação abreviada

$$h = \frac{-a_1}{2} \quad \text{e} \quad v = \frac{\sqrt{4a_2 - a_1^2}}{2}$$

as duas raízes podem ser denotadas por um par de *números complexos conjugados*:

$$r_1, r_2 = h \pm vi$$

Diz-se que essas duas raízes complexas são “conjugadas” porque elas sempre aparecem juntas, sendo que uma é a *soma* de h e vi , e a outra é a *diferença* entre h e vi . Note que elas têm o mesmo valor absoluto R .

EXEMPLO 1 Encontre as raízes da equação característica $r^2 + r + 4 = 0$. Aplicando a fórmula conhecida, temos

$$r_1, r_2 = \frac{-1 \pm \sqrt{-15}}{2} = \frac{-1 \pm \sqrt{15}\sqrt{-1}}{2} = \frac{-1}{2} \pm \frac{\sqrt{15}}{2}i$$

que constituem um par de números complexos conjugados.

Como antes, podemos usar (16.6) para verificar nossos cálculos. Se estiverem corretos, teremos $r^1 + r^2 = -a_1 (= -1)$ e $r_1 r_2 = a_2 (= 4)$. Como temos

$$\begin{aligned} r_1 + r_2 &= \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2} \right) + \left(\frac{-1}{2} - \frac{\sqrt{15}i}{2} \right) \\ &= \frac{-1}{2} + \frac{-1}{2} = -1 \end{aligned}$$

$$\begin{aligned} \text{e} \quad r_1 r_2 &= \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2} \right) \left(\frac{-1}{2} - \frac{\sqrt{15}i}{2} \right) \\ &= \left(\frac{-1}{2} \right)^2 - \left(\frac{\sqrt{15}i}{2} \right)^2 = \frac{1}{4} - \frac{-15}{4} = 4 \end{aligned}$$

nosso cálculo fica, de fato, validado.

Mesmo no caso da raiz complexa (Caso 3), podemos expressar a função complementar de uma equação diferencial de acordo com (16.7); isto é,

$$y_c = A_1 e^{(h-vi)t} + A_2 e^{(h+vi)t} = e^{ht} (A_1 e^{vit} + A_2 e^{-vit}) \quad (16.11)$$

Mas uma nova característica foi introduzida: o número i agora aparece nos expoentes das duas expressões entre parênteses. Como interpretamos tais funções exponenciais imaginárias?

Para facilitar sua interpretação, será útil primeiramente transformar essas expressões para a forma de *funções circulares* equivalentes. Como veremos daqui a em pouco, essas últimas funções envolvem, caracteristicamente, flutuações periódicas de uma variável. Por conseqüência, (16.11), a função complementar por ser traduzível para formas de função circular, também deve gerar um tipo de trajetória temporal circular.

Funções circulares

Considere um círculo cujo centro está no ponto de origem e cujo raio tem comprimento R , como mostrado na Figura 16.3. Imagine que o raio, como o ponteiro de um relógio, gire na direção anti-horária. Iniciando na posição OA , ele passa gradualmente para a posição OP , e em seguida, sucessivamente para as posições OB , OC e OD ; e, no final de um ciclo, retorna para OA . Dali em diante, o ciclo simplesmente se repetirá.

Quando estiver em uma posição específica – digamos, OP – o ponteiro do relógio formará um ângulo definido θ com a reta OA , e a extremidade do ponteiro (P) determinará uma distância vertical v e uma distância horizontal h . À medida que o ângulo θ muda durante o processo de rotação, v e h variarão, embora R não varie. Assim, as razões v/R e h/R devem variar com θ ; isto é, essas duas razões são, ambas, funções do ângulo θ . Especificamente, v/R e h/R são denominadas, respectivamente, o *seno* (a função *seno*) de θ e o *co-seno* (a função *co-seno*) de θ :

$$\text{sen } \theta \equiv \frac{v}{R} \quad (16.12)$$

$$\text{cos } \theta \equiv \frac{h}{R} \quad (16.13)$$

Em vista da conexão dessas duas funções com um círculo, elas são denominadas *funções circulares*. Contudo, uma vez que também são associadas a um triângulo, elas são denominadas, alternativamente, *funções trigonométricas*. Um outro nome (mais elegante) para elas é *funções senoidais*. As

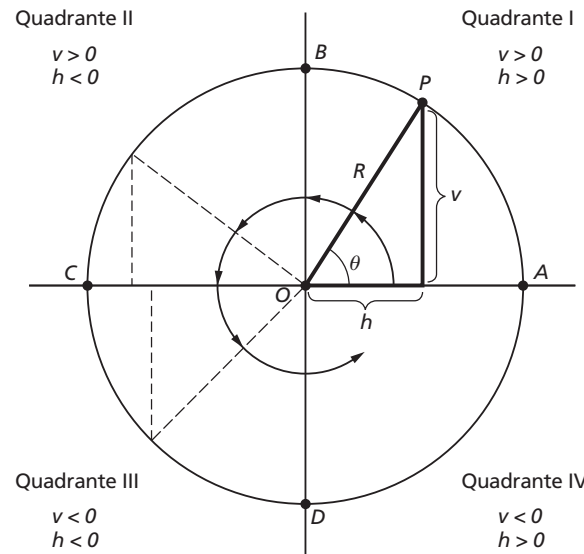


FIGURA 16.3

funções seno e co-seno não são as únicas funções circulares; outra dessas funções freqüentemente encontrada é a função *tangente*, definida como

$$\tan \theta = \frac{\sin \theta}{\cos \theta} = \frac{v}{h} \quad (h \neq 0)$$

Contudo, aqui, nossa maior preocupação será com as funções seno e co-seno.

A variável independente em uma função circular é o ângulo θ , de modo que o mapeamento envolvido aqui é de um *ângulo* para uma *razão entre duas distâncias*. Ângulos geralmente são medidos em *graus* (por exemplo, 30, 45 e 90°); no trabalho analítico, todavia, é mais conveniente medir ângulos em *radianos*. A vantagem da medida em radianos origina-se do fato de que, quando θ é medido dessa maneira, as derivadas de funções circulares resultarão em expressões mais elegantes – assim como a base e nos dá derivadas mais elegantes para funções exponenciais e logarítmicas. Mas exatamente quanto é um radiano? Para explicar isso, vamos voltar à Figura 16.3, onde desenhamos um ponto P de modo que o comprimento do *arco* AP é exatamente igual ao do raio R . Então, um *radiano* (abreviado por *rad*) pode ser definido como o tamanho do ângulo θ (na Figura 16.3) formado por tal arco de comprimento R . Uma vez que a circunferência do círculo tem um comprimento total de $2\pi R$ (onde $\pi = 3,14159\dots$), um círculo completo deve envolver um ângulo de 2π rad no total. Em termos de graus, entretanto, um círculo completo faz um ângulo de 360°; assim, igualando 360° a 2π rad, podemos chegar à seguinte tabela de conversão:

Graus	360	270	180	90	45	0
Radianos	2π	$\frac{3\pi}{2}$	π	$\frac{\pi}{2}$	$\frac{\pi}{4}$	0

Propriedades das funções seno e co-seno

Dado o comprimento de R , o valor de $\sin \theta$ depende do modo como o valor v varia em resposta às variações no ângulo θ . Na posição inicial OA , temos $v = 0$. À medida que o ponteiro se move em sentido anti-horário, v começa a assumir um valor positivo crescente, culminando no valor máximo de $v = R$ quando o ponteiro coincidir com OB , isto é, quando $\theta = \pi/2$ rad ($= 90^\circ$). Continuando o movimento, v ficará gradualmente mais curto, até que seu valor se tornará zero quando o ponteiro estiver na posição OC , isto é, quando $\theta = \pi$ rad ($= 180^\circ$). À medida que o ponteiro entra no terceiro quadrante, v começa a assumir valores negativos; na posição OD , temos $v = -R$. No quarto quadrante, v ainda é negativo, mas crescerá do valor de $-R$ para o valor de $v = 0$, que é atingido quando o ponteiro volta a OA – isto é, quando $\theta = 2\pi$ rad ($= 360^\circ$). Então o ciclo se repete.

Quando esses valores ilustrativos de v são substituídos em (16.12), podemos obter os resultados mostrados na linha “ $\sin \theta$ ” da Tabela 16.1. Contudo, se quiser uma descrição mais completa da função seno, veja a Figura 16.4a, que mostra o gráfico dos valores de $\sin \theta$ em função de θ (expressos em radianos).

O valor de $\cos \theta$, por outro lado, depende do modo como h varia em resposta a variações em θ . Na posição inicial OA , temos $h = R$. Então h diminui gradualmente até $h = 0$ quando $\theta = \pi/2$ (posição OB). No segundo quadrante, h fica negativo, e quando $\theta = \pi$ (posição OC), $h = -R$. O valor de h aumenta gradualmente de $-R$ até zero no terceiro quadrante e, quando $\theta = 3\pi/2$ (posição OD), constatamos que $h = 0$. No quarto quadrante, h fica positivo novamente e, quando o ponteiro volta à posição OA ($\theta = 2\pi$), temos novamente $h = R$. Então o ciclo se repete.

A substituição desses valores ilustrativos de h em (16.13) dá os resultados que estão na última linha da Tabela 16.1, mas a Figura 16.4b dá uma descrição mais completa da função co-seno.

As funções $\sin \theta$ e $\cos \theta$ compartilham o mesmo domínio, a saber, o conjunto de todos os números reais (as medidas de θ em radianos). Por esse critério, podemos mostrar que um ângulo *negativo* refere-se simplesmente à rotação do ponteiro de relógio em sentido inverso; por exemplo, um movimento no sentido horário de OA para OD na Figura 16.3 gera um ângulo de $-\pi/2$ rad ($= -90^\circ$). As duas funções também têm a mesma imagem, a saber, o intervalo fechado $[-1, 1]$. Por essa razão, na Figura 16.4, os gráficos de $\sin \theta$ e $\cos \theta$ estão confinados a uma faixa horizontal definida.

TABELA 16.1

θ	0	$\frac{1}{2}\pi$	π	$\frac{3}{2}\pi$	2π
sen θ	0	1	0	-1	0
cos θ	1	0	-1	0	1

Uma importante propriedade das funções seno e co-seno é que ambas são *periódicas*; seus valores se repetirão a cada 2π rad (um círculo completo) pelo qual passar o ângulo θ . Por conseguinte, diz-se que cada função tem um *período* de 2π . Em vista dessa característica de periodicidade, as seguintes equações são válidas (para qualquer inteiro n):

$$\text{sen}(\theta + 2n\pi) = \text{sen } \theta \quad \cos(\theta + 2n\pi) = \cos \theta$$

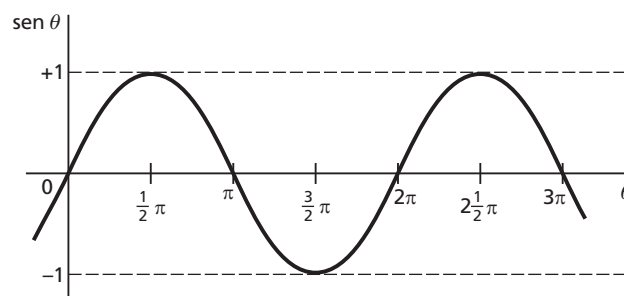
Isto é, somar (ou subtrair) qualquer múltiplo inteiro de 2π a (ou de) qualquer ângulo θ não afetará nem o valor de $\text{sen } \theta$ nem o de $\cos \theta$.

Os gráficos das funções seno e co-seno indicam uma faixa constante de flutuação em cada período, a saber, ± 1 , o que às vezes é descrito alternativamente dizendo-se que a *amplitude* de flutuação é 1. Em virtude do período idêntico e da amplitude idêntica, vemos que a curva $\cos \theta$, se deslocada para a direita por $\pi/2$, coincidirá exatamente com a curva $\text{sen } \theta$. Por conseguinte, diz-se que essas duas curvas são diferentes apenas em *fase*, isto é, apenas na localização do pico em cada período. Esse fato pode ser enunciado simbolicamente pela equação

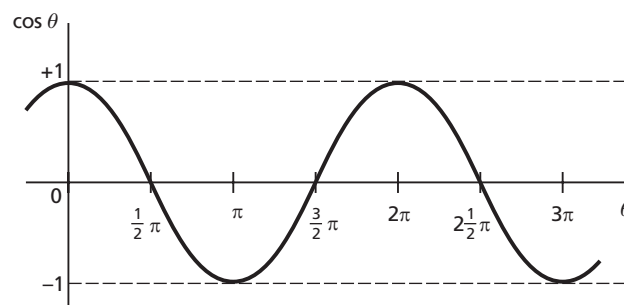
$$\cos \theta = \text{sen}\left(\theta + \frac{\pi}{2}\right)$$

As funções seno e co-seno obedecem a certas identidades. Entre essas, as mais frequentemente usadas são

$$\begin{aligned} \text{sen}(-\theta) &\equiv -\text{sen } \theta \\ \cos(-\theta) &\equiv \cos \theta \end{aligned} \quad (16.14)$$



(a)



(b)

FIGURA 16.4

$$\sin^2 \theta + \cos^2 \theta \equiv 1 \quad [\text{onde } \sin^2 \theta \equiv (\sin \theta)^2 \text{ etc.}] \quad (16.15)$$

$$\begin{aligned} \sin(\theta_1 \pm \theta_2) &\equiv \sin \theta_1 \cos \theta_2 \pm \cos \theta_1 \sin \theta_2 \\ \cos(\theta_1 \mp \theta_2) &\equiv \cos \theta_1 \cos \theta_2 \pm \sin \theta_1 \sin \theta_2 \end{aligned} \quad (16.16)$$

O par de identidades (16.14) indica que a função co-seno é simétrica em relação ao eixo vertical (isto é, θ e $-\theta$ sempre resultam no mesmo valor de co-seno), enquanto a função seno não é. A expressão (16.15) mostra o fato de que, para qualquer grandeza de θ , a soma dos quadrados de seus seno e co-seno é sempre a unidade. E o conjunto de identidades em (16.16) dá o seno e o co-seno da soma e da diferença de dois ângulos θ_1 e θ_2 .

Por fim, uma palavra sobre derivadas. Como as funções $\sin \theta$ e $\cos \theta$ são contínuas e suaves, ambas são diferenciáveis. As derivadas, $d(\sin \theta)/d\theta$ e $d(\cos \theta)/d\theta$, podem ser obtidas tomando os limites, respectivamente, dos quocientes de diferenças $\Delta(\sin \theta)/\Delta\theta$ e $\Delta(\cos \theta)/\Delta\theta$ à medida que $\Delta\theta \rightarrow 0$.

Os resultados, apresentados aqui sem demonstração, são

$$\frac{d}{d\theta} \sin \theta = \cos \theta \quad (16.17)$$

$$\frac{d}{d\theta} \cos \theta = -\sin \theta \quad (16.18)$$

Entretanto, devemos enfatizar que essas fórmulas de derivadas são válidas somente quando θ é medido em radianos; se for medido em graus, por exemplo, (16.17) se tornará, por sua vez, $d(\sin \theta)/d\theta = (\pi/180) \cos \theta$. É com o intuito de nos livrarmos do fator $(\pi/180)$ que, no trabalho analítico, as medidas em radianos são preferidas às medidas em graus.

EXEMPLO 2

Encontre a inclinação da curva $\sin \theta$ em $\theta = \pi/2$. A inclinação da curva do seno é dada por sua derivada ($= \cos \theta$). Assim, em $\theta = \pi/2$, a inclinação deve ser $\cos(\pi/2) = 0$. Você pode consultar a Figura 16.4 para verificar esse resultado.

EXEMPLO 3

Encontre a derivada segunda de $\sin \theta$. Por (16.17), sabemos que a derivada primeira de $\sin \theta$ é $\cos \theta$, por conseguinte, a derivada segunda desejada é

$$\frac{d^2}{d\theta^2} \sin \theta = \frac{d}{d\theta} \cos \theta = -\sin \theta$$

Relações de Euler

Na Seção 9.5, demonstramos que qualquer função que tenha derivadas finitas e contínuas até a ordem desejada pode ser expandida em uma função polinomial. Além do mais, se o resto R_n da série de Taylor (expansão em qualquer ponto x_0) ou da série de Maclaurin (expansão em $x_0 = 0$) resultante acaso se aproxime de zero à medida que o número n de termos se torna infinito, o polinômio pode ser escrito como uma série infinita. Agora, vamos expandir as funções seno e co-seno e então tentar mostrar como as expressões exponenciais imaginárias encontradas em (16.11) podem ser transformadas em funções circulares que têm expansões equivalentes.

Para a função seno, escreva $\phi(\theta) = \sin \theta$; segue-se, então, que $\phi(0) = \sin 0 = 0$. Por derivação sucessiva, podemos obter

$$\left. \begin{aligned} \phi'(\theta) &= \cos \theta \\ \phi''(\theta) &= -\sin \theta \\ \phi'''(\theta) &= -\cos \theta \\ \phi^{(4)}(\theta) &= \sin \theta \\ \phi^{(5)}(\theta) &= \cos \theta \\ &\vdots \end{aligned} \right\} \Rightarrow \left\{ \begin{aligned} \phi'(0) &= \cos 0 = 1 \\ \phi''(0) &= -\sin 0 = 0 \\ \phi'''(0) &= -\cos 0 = -1 \\ \phi^{(4)}(0) &= \sin 0 = 0 \\ \phi^{(5)}(0) &= \cos 0 = 1 \\ &\vdots \end{aligned} \right.$$

Quando substituídas em (9.14), onde θ agora está no lugar de x , teremos a seguinte série de Maclaurin com resto:

$$\sin \theta = 0 + \theta + 0 - \frac{\theta^3}{3!} + 0 + \frac{\theta^5}{5!} + \cdots + \frac{\phi^{(n+1)}(p)}{(n+1)!} \theta^{n+1}$$

Agora, a expressão $\phi^{(n+1)}(p)$ no último termo (resto), que representa a derivada de ordem $(n+1)$ calculada em $\theta = p$, pode assumir apenas a forma de $\pm \cos p$ ou $\pm \sin p$ e, como tal, pode assumir apenas um valor no intervalo $[-1, 1]$, independentemente do tamanho de n . Por outro lado, $(n+1)!$ crescerá rapidamente à medida que $n \rightarrow \infty$ – de fato, muito mais rapidamente do que θ^{n+1} à medida que n aumenta. Por conseguinte, o resto se aproximará de zero à medida que $n \rightarrow \infty$ e, portanto, podemos expressar a série de Maclaurin como uma série infinita:

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \quad (16.19)$$

De modo semelhante, se escrevermos $\psi(\theta) = \cos \theta$, então $\psi(0) = \cos 0 = 1$, e as derivadas sucessivas serão

$$\left. \begin{array}{l} \psi'(\theta) = -\sin \theta \\ \psi''(\theta) = -\cos \theta \\ \psi'''(\theta) = \sin \theta \\ \psi^{(4)}(\theta) = \cos \theta \\ \psi^{(5)}(\theta) = -\sin \theta \\ \vdots \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \psi'(0) = -\sin 0 = 0 \\ \psi''(0) = -\cos 0 = -1 \\ \psi'''(0) = \sin 0 = 0 \\ \psi^{(4)}(0) = \cos 0 = 1 \\ \psi^{(5)}(0) = -\sin 0 = 0 \\ \vdots \end{array} \right.$$

Com base nessas derivadas, podemos expandir $\cos \theta$ da seguinte maneira:

$$\cos \theta = 1 + 0 - \frac{\theta^2}{2!} + 0 + \frac{\theta^4}{4!} + \cdots + \frac{\psi^{(n+1)}(p)}{(n+1)!} \theta^{n+1}$$

Visto que o resto novamente tenderá a zero à medida que $n \rightarrow \infty$, a função co-seno também pode ser expressa como uma série infinita, como se segue:

$$\cos \theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \quad (16.20)$$

Você deve ter notado que, com (16.19) e (16.20) à mão, agora podemos construir uma tabela de valores de senos e co-senos para todos os valores possíveis de θ (em radianos). Entretanto, nosso interesse imediato é achar a relação entre expressões exponenciais imaginárias e funções circulares. Com essa finalidade, agora vamos expandir as duas expressões exponenciais $e^{j\theta}$ e $e^{-j\theta}$. O leitor perceberá que essas expressões não são nada mais do que casos especiais da expressão e^x , que já demonstramos, em (10.6), que têm a expansão

$$e^x = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \frac{1}{4!} x^4 + \cdots$$

Por conseguinte, fazendo $x = i\theta$, podemos obter imediatamente

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \dots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \dots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots\right) \end{aligned}$$

De modo semelhante, fazendo $x = -i\theta$, teremos o seguinte resultado:

$$\begin{aligned} e^{-i\theta} &= 1 - i\theta + \frac{(-i\theta)^2}{2!} + \frac{(-i\theta)^3}{3!} + \frac{(-i\theta)^4}{4!} + \frac{(-i\theta)^5}{5!} + \dots \\ &= 1 - i\theta - \frac{\theta^2}{2!} + \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} - \frac{i\theta^5}{5!} - \dots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots\right) - i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots\right) \end{aligned}$$

Substituindo (16.19) e (16.20) nesses dois resultados, podemos estabelecer, de imediato, os seguintes pares de identidades – conhecidos como *relações de Euler*:

$$e^{i\theta} \equiv \cos \theta + i \sin \theta \quad (16.21)$$

$$e^{-i\theta} \equiv \cos \theta - i \sin \theta \quad (16.21')$$

Essas expressões nos habilitarão a traduzir qualquer função exponencial imaginária em uma combinação linear equivalente de funções seno e co-seno, e vice-versa.

EXEMPLO 4 Encontre o valor de $e^{i\pi}$. Primeiro, vamos converter essa expressão em uma expressão trigonométrica. Fazendo $\theta = \pi$ em (16.21), constatamos que $e^{i\pi} = \cos \pi + i \sin \pi$. Visto que $\cos \pi = -1$ e $\sin \pi = 0$, deduz-se que $e^{i\pi} = -1$.

EXEMPLO 5 Mostre que $e^{-i\pi/2} = -i$. Fazendo $\theta = \pi/2$ em (16.21'), temos

$$e^{-i\pi/2} = \cos \frac{\pi}{2} - i \sin \frac{\pi}{2} = 0 - i(1) = -i$$

Representações alternativas de números complexos

Até aqui, representamos um par de números complexos conjugados na forma geral $(h \pm vi)$. Uma vez que h e v referem-se à abscissa e à ordenada do sistema de coordenadas cartesianas de um diagrama de Argand, a expressão $(h \pm vi)$ representa a *forma cartesiana* de um par de números complexos conjugados. Como um subproduto da discussão das funções circulares e das relações de Euler, agora podemos expressar $(h \pm vi)$ de dois outros modos.

Referindo-nos à Figura 16.2, vemos que, tão logo h e v sejam especificados, o ângulo θ e o valor de R também tornam-se determinados. Uma vez que um dado θ e um dado R juntos podem identificar um único ponto no diagrama de Argand, podemos empregar θ e R para especificar o

par específico de números complexos. Reescrevendo as definições das funções seno e co-seno em (16.12) e (16.13) como

$$v = R \operatorname{sen} \theta \quad \text{e} \quad h = R \cos \theta \quad (16.22)$$

os números complexos conjugados $(h \pm vi)$ podem ser transformados como se segue:

$$h \pm vi = R \cos \theta \pm Ri \operatorname{sen} \theta = R(\cos \theta \pm i \operatorname{sen} \theta)$$

Fazendo isso, na verdade passamos de coordenadas cartesianas dos números complexos $(h \text{ e } v)$ para o que denominamos suas *coordenadas polares* $(R \text{ e } \theta)$. De acordo com isso, a expressão do lado direito da equação precedente exemplifica a *forma polar* de um par de números complexos conjugados.

Além do mais, em vista das relações de Euler, a forma polar também pode ser reescrita na *forma exponencial*, como se segue: $R(\cos \theta \pm i \operatorname{sen} \theta) = Re^{\pm i\theta}$. Daí, temos um total de três representações alternativas dos números complexos conjugados:

$$h \pm vi = R(\cos \theta \pm i \operatorname{sen} \theta) = Re^{\pm i\theta} \quad (16.23)$$

Se tivermos os valores de R e θ , a transformação para h e v é direta: usamos as duas equações em (16.22). E a transformação inversa? Tendo valores dados de h e v , não surge nenhuma dificuldade para achar o valor correspondente de R , que é igual a $\sqrt{h^2 + v^2}$. Mas surge uma ligeira ambigüidade em relação a θ : o valor desejado de θ (em radianos) é que satisfaz as duas condições $\cos \theta = h/R$ e $\operatorname{sen} \theta = v/R$; mas, para valores dados de h e v , θ não é único! (Por quê?) Felizmente, o problema não é sério pois, confinando nossa atenção ao intervalo $[0, 2\pi)$ no domínio, a indeterminação é rapidamente resolvida.

EXEMPLO 6

Encontre a forma cartesiana do número complexo $5e^{3i\pi/2}$. Aqui, temos $R = 5$ e $\theta = 3\pi/2$; portanto, por (16.22) e pela Tabela 16.1,

$$h = 5 \cos \frac{3\pi}{2} = 0 \quad \text{e} \quad v = 5 \operatorname{sen} \frac{3\pi}{2} = -5$$

Assim, a forma cartesiana é simplesmente $h + vi = -5i$.

EXEMPLO 7

Encontre as formas polar e exponencial de $(1 + \sqrt{3}i)$. Neste caso, temos $h = 1$ e $v = \sqrt{3}$; assim, $R = \sqrt{1 + 3} = 2$. A Tabela 16.1 não tem nenhuma utilidade para localizar o valor de θ desta vez, mas a Tabela 16.2, que relaciona alguns valores selecionados de $\operatorname{sen} \theta$ e $\cos \theta$, ajudará. Especificamente, estamos buscando o valor de θ tal que $\cos \theta = h/R = 1/2$ e $\operatorname{sen} \theta = v/R = \sqrt{3}/2$. O valor $\theta = \pi/3$ cumpre os requisitos. Assim, de acordo com (16.23), a transformação desejada é

$$1 + \sqrt{3}i = 2 \left(\cos \frac{\pi}{3} + i \operatorname{sen} \frac{\pi}{3} \right) = 2e^{i\pi/3}$$

θ	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{3\pi}{4}$
$\operatorname{sen} \theta$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}} \left(= \frac{\sqrt{2}}{2} \right)$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}} \left(= \frac{\sqrt{2}}{2} \right)$
$\cos \theta$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}} \left(= \frac{\sqrt{2}}{2} \right)$	$\frac{1}{2}$	$\frac{-1}{\sqrt{2}} \left(= -\frac{\sqrt{2}}{2} \right)$

TABELA 16.2

Antes de encerrar este tópico, vamos observar uma extensão importante do resultado em (16.23). Supondo que temos a n -ésima potência de um número complexo – digamos, $(h + vi)^n$ –, como escrevemos suas formas polar e exponencial? A forma exponencial é mais fácil de deduzir. Uma vez que $h + vi = Re^{i\theta}$, segue-se que

$$(h + vi)^n = (Re^{i\theta})^n = R^n e^{in\theta}$$

De modo semelhante, podemos escrever

$$(h - vi)^n = (Re^{-i\theta})^n = R^n e^{-in\theta}$$

Note que a potência n provocou duas mudanças: (1) R agora se torna R^n ; e (2) θ agora se torna $n\theta$. Quando essas duas mudanças são inseridas na forma polar (16.23), constatamos que

$$(h \pm vi)^n = R^n (\cos n\theta \pm i \sin n\theta) \quad (16.23')$$

Isto é,

$$[R(\cos \theta \pm i \sin \theta)]^n = R^n (\cos n\theta \pm i \sin n\theta)$$

Conhecido como *teorema de De Moivre*, esse resultado indica que, para elevar um número complexo à n -ésima potência, devemos simplesmente modificar suas coordenadas polares elevando R à n -ésima potência e então multiplicando θ por n .

EXERCÍCIO 16.2

- Encotre as raízes das seguintes equações quadráticas:

(a) $r^2 - 3r + 9 = 0$	(c) $2x^2 + x + 8 = 0$
(b) $r^2 + 2r + 17 = 0$	(d) $2x^2 - x + 1 = 0$
- (a) Quantos graus há em um radiano?
(b) Quantos radianos há em um grau?
- Com referência à Figura 16.3, e usando o teorema de Pitágoras, prove que

(a) $\sin^2 \theta + \cos^2 \theta \equiv 1$	(b) $\sin \frac{\pi}{4} = \cos \frac{\pi}{4} = \frac{1}{\sqrt{2}}$
--	--
- Por meio das identidades (16.14), (16.15) e (16.16), mostre que:

(a) $\sin 2\theta \equiv 2 \sin \theta \cos \theta$	
(b) $\cos 2\theta \equiv 1 - 2 \sin^2 \theta$	
(c) $\sin(\theta_1 + \theta_2) + \sin(\theta_1 - \theta_2) \equiv 2 \sin \theta_1 \cos \theta_2$	
(d) $1 + \tan^2 \theta \equiv \frac{1}{\cos^2 \theta}$	
(e) $\sin\left(\frac{\pi}{2} - \theta\right) \equiv \cos \theta$	(f) $\cos\left(\frac{\pi}{2} - \theta\right) \equiv \sin \theta$
- Aplicando a regra da cadeia:

(a) Escreva as fórmulas de derivadas para $\frac{d}{d\theta} \sin f(\theta)$ e $\frac{d}{d\theta} \cos f(\theta)$, onde $f(\theta)$ é uma função de θ .
(b) Encontre as derivadas de $\cos \theta^3$, $\sin(\theta^2 + 3\theta)$, $\cos e^\theta$ e $\sin(1/\theta)$.
- Pelas relações de Euler, deduza que:

(a) $e^{-i\pi} = -1$	(c) $e^{i\pi/4} = \frac{\sqrt{2}}{2}(1+i)$
(b) $e^{i\pi/3} = \frac{1}{2}(1+\sqrt{3}i)$	(d) $e^{-3i\pi/4} = -\frac{\sqrt{2}}{2}(1+i)$

7. Encontre a forma cartesiana de cada número complexo:

$$(a) 2 \left(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right) \quad (b) 4e^{i\pi/3} \quad (c) \sqrt{2}e^{-i\pi/4}$$

8. Encontre as formas polar e exponencial dos seguintes números complexos:

$$(a) \frac{3}{2} + \frac{3\sqrt{3}}{2}i \quad (b) 4(\sqrt{3} + i)$$

16.3 Análise do caso da raiz complexa

Tendo à disposição os conceitos de números complexos e funções circulares, agora estamos preparados para abordar o caso da raiz complexa (Caso 3) a que nos referimos na Seção 16.1. Lembre-se que a classificação dos três casos segundo a natureza das raízes características preocupa-se somente com a função complementar de uma equação diferencial. Assim, podemos continuar concentrando nossa atenção na equação homogênea

$$y''(t) + a_1 y'(t) + a_2 y = 0 \quad [\text{reproduzida de (16.4)}]$$

A função complementar

Quando os valores dos coeficientes a_1 e a_2 são tais que $a_1^2 < 4a_2$, as raízes características serão o par de números complexos conjugados

$$r_1, r_2 = h \pm vi$$

$$\text{onde} \quad h = -\frac{1}{2}a_1 \quad \text{e} \quad v = \frac{1}{2}\sqrt{4a_2 - a_1^2}$$

Assim, a função complementar, como já havia sido previsto, terá a forma

$$y_c = e^{ht} (A_1 e^{vit} + A_2 e^{-vit}) \quad [\text{reproduzida de (16.11)}]$$

Em primeiro lugar, vamos transformar as expressões exponenciais imaginárias entre parênteses em expressões trigonométricas equivalentes, de modo que possamos interpretar a função complementar como uma função circular. Isso pode ser feito usando as relações de Euler. Fazendo $\theta = vt$ em (16.21) e (16.21'), constatamos que

$$e^{vit} = \cos vt + i \sin vt \quad \text{e} \quad e^{-vit} = \cos vt - i \sin vt$$

Dessas expressões, segue-se que a função complementar em (16.11) pode ser reescrita como

$$\begin{aligned} y_c &= e^{ht} [A_1 (\cos vt + i \sin vt) + A_2 (\cos vt - i \sin vt)] \\ &= e^{ht} [(A_1 + A_2) \cos vt + (A_1 - A_2) i \sin vt] \end{aligned} \quad (16.24)$$

Além do mais, se empregarmos os símbolos abreviados

$$A_5 \equiv A_1 + A_2 \quad \text{e} \quad A_6 \equiv (A_1 - A_2) i$$

é possível simplificar (16.24) para[†]

$$y_c = e^{ht} (A_5 \cos vt + A_6 \sin vt) \quad (16.24')$$

onde as novas constantes arbitrárias A_5 e A_6 serão definidas mais adiante.

Se você for metodoso, talvez se sinta um tanto inseguro com a substituição de θ por vt no procedimento precedente. A variável θ mede um ângulo, mas vt é uma grandeza em unidades de t (em nosso contexto, tempo). Por conseguinte, como podemos fazer a substituição $\theta = vt$? A resposta a essa pergunta pode ser melhor explicada com referência ao *círculo unitário* (um círculo de raio $R = 1$) na Figura 16.5. É verdade que temos usado θ para designar um ângulo; mas, visto que o ângulo é medido em unidades de radianos, o valor de θ é sempre a razão entre o comprimento do arco AB e o raio R . Quando $R = 1$, temos, especificamente,

$$\theta \equiv \frac{\text{arco } AB}{R} \equiv \frac{\text{arco } AB}{1} \equiv \text{arco } AB$$

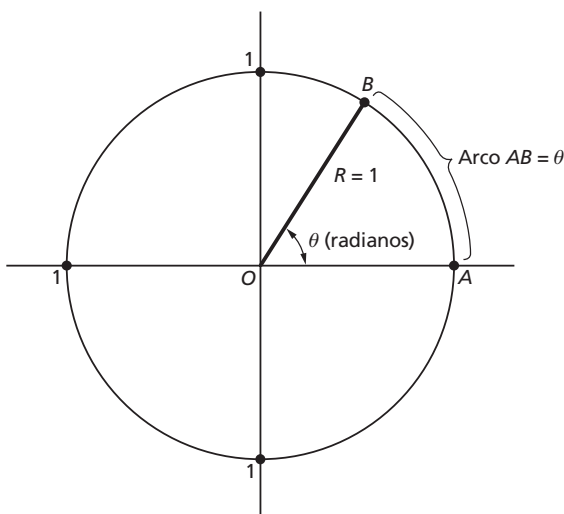
Em outras palavras, θ não é somente a medida do ângulo em radianos, mas também o comprimento do arco AB , que é um número, e não um ângulo. Se a passagem do tempo for registrada na circunferência do círculo unitário (em sentido anti-horário), e não sobre uma linha reta como fazemos quando queremos representar graficamente uma série temporal, na verdade não faz nenhuma diferença se considerarmos o lapso de tempo como um aumento na medida do ângulo θ em radianos ou como um aumento no comprimento do arco AB . Além disso, ainda que $R \neq 1$, a mesma linha de raciocínio pode ser aplicada, exceto que, nesse caso, θ será igual a $(\text{arco } AB)/R$; isto é, o ângulo θ e o arco AB guardarão uma proporção fixa entre si, em vez de serem iguais. Assim, a substituição $\theta = vt$ é, de fato, legítima.

Um exemplo de solução

Vamos achar a solução da equação diferencial

$$y''(t) + 2y'(t) + 17y = 34$$

FIGURA 16.5



[†] O fato de que, ao definir A_6 , nela incluímos o número imaginário i não é, de modo algum, uma tentativa de "varrer a sujeira para baixo do tapete". Como A_6 é uma constante arbitrária, ela pode assumir um valor imaginário, bem como um valor real. E tampouco é verdade que, do modo como foi definida, A_6 necessariamente resultará em um número imaginário. Na verdade, se A_1 e A_2 formarem um par de números complexos conjugados, digamos, $m \pm ni$, então A_5 e A_6 serão ambas reais: $A_5 = A_1 + A_2 = (m + ni) + (m - ni) = 2m$ e $A_6 = (A_1 - A_2)i = [(m + ni) - (m - ni)]i = (2ni)i = -2n$.

com as condições iniciais $y(0) = 3$ e $y'(0) = 11$.

Uma vez que $a_1 = 2$, $a_2 = 17$, e $b = 34$, podemos constatar, de imediato, que a solução particular é

$$y_p = \frac{b}{a_2} = \frac{34}{17} = 2 \quad [\text{por (16.3)}]$$

Além do mais, visto que $a_1^2 = 4 < 4a_2 = 68$, as raízes características serão o par de números complexos conjugados $(h \pm vi)$, onde

$$h = -\frac{1}{2}a_1 = -1 \quad \text{e} \quad v = \frac{1}{2}\sqrt{4a_2 - a_1^2} = \frac{1}{2}\sqrt{64} = 4$$

Daí, por (16.24'), a função complementar é

$$y_c = e^{-t} (A_5 \cos 4t + A_6 \sin 4t)$$

Combinando y_c e y_p , a solução geral pode ser expressa como

$$y(t) = e^{-t} (A_5 \cos 4t + A_6 \sin 4t) + 2$$

Para definir as constantes A_5 e A_6 , utilizamos as duas condições iniciais. Primeiro, fazendo $t = 0$ na solução geral, constatamos que

$$\begin{aligned} y(0) &= e^0 (A_5 \cos 0 + A_6 \sin 0) + 2 \\ &= (A_5 + 0) + 2 = A_5 + 2 \quad [\cos 0 = 1; \sin 0 = 0] \end{aligned}$$

Assim, pela condição inicial $y(0) = 3$, podemos especificar $A_5 = 1$. Em seguida, vamos diferenciar a solução geral em relação a t — usando a regra do produto e as fórmulas de derivada (16.17) e (16.18) e tendo em mente, ao mesmo tempo, a regra da cadeia [Exercício 16.2–5] — para achar $y'(t)$ e então $y'(0)$:

$$y'(t) = -e^{-t} (A_5 \cos 4t + A_6 \sin 4t) + e^{-t} [-4A_5 \sin 4t + 4A_6 \cos 4t]$$

de modo que

$$\begin{aligned} y'(0) &= -(A_5 \cos 0 + A_6 \sin 0) + (-4A_5 \sin 0 + 4A_6 \cos 0) \\ &= -(A_5 + 0) + (0 + 4A_6) = 4A_6 - A_5 \end{aligned}$$

Pela segunda condição inicial $y'(0) = 11$, e tendo em vista que $A_5 = 1$, fica claro que $A_6 = 3$.[†] Portanto, a solução definida é

$$y(t) = e^{-t} (\cos 4t + 3 \sin 4t) + 2 \quad (16.25)$$

Como antes, o componente $y_p (= 2)$ pode ser interpretado como o nível de equilíbrio intertemporal de y , ao passo que o componente y_c representa o desvio em relação ao equilíbrio. Por causa da presença de funções circulares em y_c , podemos esperar que a trajetória temporal (16.25) exiba um padrão flutuante. Mas qual padrão específico ela envolverá?

[†] Note que, aqui, A_6 realmente é um número real, embora tenhamos incluído o número imaginário i em sua definição.

A trajetória temporal

Já estamos familiarizados com as trajetórias de uma função seno ou co-seno simples, como mostrado na Figura 16.4. Agora, devemos estudar as trajetórias de certas variantes e combinações de funções seno e co-seno de modo que possamos interpretar, no geral, a função complementar (16.24')

$$y_c = e^{ht} (A_5 \cos vt + A_6 \sin vt)$$

e, em particular, o componente y_c de (16.25).

Vamos examinar, em primeiro lugar, o termo $(A_5 \cos vt)$. Por si só, a expressão $(\cos vt)$ é uma função circular de (vt) , com período 2π ($= 6,2832$) e amplitude 1. O período 2π significa que o gráfico repetirá sua configuração toda vez que (vt) aumentar de 2π . Quando apenas t é tomada como a variável independente, contudo, a repetição ocorrerá toda vez que t aumentar de $2\pi/v$, de modo que, com referência a t — como é adequado em análise econômica dinâmica —, consideraremos que o período de $(\cos vt)$ é $2\pi/v$. (Contudo, a amplitude continua sendo 1.) Agora, quando anexarmos uma constante multiplicativa A_5 a $(\cos vt)$, ela fará com que a faixa de flutuação mude de ± 1 para $\pm A_5$. Assim, a amplitude agora se torna A_5 , embora o período continue não sendo afetado pela constante. Em suma, $(A_5 \cos vt)$ é uma função co-seno de t , com período $2\pi/v$ e amplitude A_5 . Pelo mesmo critério, $(A_6 \sin vt)$ é uma função seno de t , com período $2\pi/v$ e amplitude A_6 .

Como há um período comum, a soma $(A_5 \cos vt + A_6 \sin vt)$ também apresentará um ciclo repetitivo toda vez que t aumentar de $2\pi/v$. Para mostrar isso com maior rigor, vamos notar que, para valores de A_5 e A_6 dados, sempre podemos achar duas constantes A e ε , tais que

$$A_5 = A \cos \varepsilon \quad \text{e} \quad A_6 = -A \sin \varepsilon$$

Assim, podemos expressar a soma citada como

$$\begin{aligned} A_5 \cos vt + A_6 \sin vt &= A \cos \varepsilon \cos vt - A \sin \varepsilon \sin vt \\ &= A(\cos vt \cos \varepsilon - \sin vt \sin \varepsilon) \\ &= A \cos(vt + \varepsilon) \quad [\text{por 16.16}] \end{aligned}$$

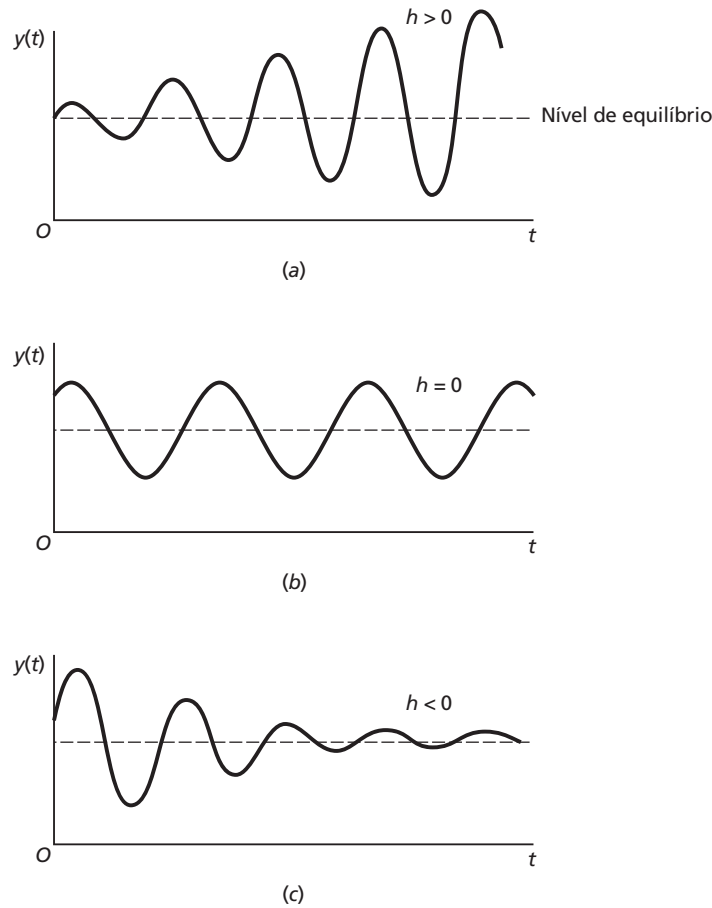
Essa é uma função co-seno modificada de t , com amplitude A e período $2\pi/v$ porque, toda vez que t aumentar de $2\pi/v$, $(vt + \varepsilon)$ aumentará de 2π , o que completará um ciclo na curva do co-seno.

Se y_c consistisse apenas na expressão $(A_5 \cos vt + A_6 \sin vt)$, a implicação seria que a trajetória temporal de y seria uma flutuação sem fim, de amplitude constante, ao redor do valor de equilíbrio de y , tal como representada por y_p . Mas, na verdade, também há o termo multiplicativo e^{ht} a considerar. Este último termo é de grande importância pois, como veremos, é a chave para determinar se a trajetória temporal convergirá ou não.

Se $h > 0$, o valor de e^{ht} aumentará continuamente à medida que t aumentar. Isso produzirá um efeito de aumento da amplitude de $(A_5 \cos vt + A_6 \sin vt)$ e causará desvios cada vez maiores em relação ao equilíbrio em cada ciclo sucessivo. Como ilustrado na Figura 16.6a, nesse caso a trajetória temporal será caracterizada por *flutuação explosiva*. Se, por outro lado, $h = 0$, então $e^{ht} = 1$ e a função complementar será simplesmente $(A_5 \cos vt + A_6 \sin vt)$, que já demonstramos ter uma amplitude constante. Neste segundo caso, cada ciclo apresentará um padrão uniforme de desvio em relação ao equilíbrio, como ilustrado pela trajetória temporal na Figura 16.6b. Essa é uma trajetória temporal com *flutuação uniforme*. Por fim, se $h < 0$, o termo e^{ht} decrescerá continuamente à medida que t aumentar, e cada ciclo sucessivo terá uma amplitude menor que o precedente, de um modo muito parecido ao de uma ondulação na superfície da água que aos poucos se desfaz. Esse caso é ilustrado na Figura 16.6c, onde a trajetória temporal é caracterizada por *flutuação amortecida*. A solução em (16.25), com $h = -1$, exemplifica este último caso. Deve ficar bem claro que somente o caso de flutuação amortecida pode produzir uma trajetória temporal *convergente*; nos outros dois casos, a trajetória temporal é *não-convergente* ou *divergente*.[†]

[†] Vamos usar as duas palavras *não-convergente* e *divergente* indiferentemente, embora a última seja mais estritamente aplicável à variedade de não-convergência explosiva do que à uniforme.

FIGURA 16.6



Em todos os três diagramas da Figura 16.6, o equilíbrio intertemporal é admitido como constante. Se for um equilíbrio móvel, os três tipos de trajetória temporal retratados ainda flutuam ao redor dele, mas, uma vez que a representação gráfica de um equilíbrio móvel é geralmente uma curva, e não uma reta horizontal, a flutuação assumirá a natureza de, digamos, uma série de ciclos de negócios ao redor de uma tendência secular.

A estabilidade dinâmica do equilíbrio

O conceito de convergência da trajetória temporal de uma variável está vinculado inextricavelmente ao conceito de estabilidade dinâmica do equilíbrio intertemporal daquela variável. Especificamente, o equilíbrio é dinamicamente estável se, e somente se, a trajetória temporal for convergente. Portanto, a condição de convergência da trajetória $y(t)$, a saber, $h < 0$ (Figura 16.6c) é também a condição para estabilidade dinâmica do equilíbrio intertemporal de y .

Você recordará que, para os Casos 1 e 2, onde as raízes características são reais, a condição para estabilidade dinâmica de equilíbrio é que toda raiz característica seja negativa. No presente caso (Caso 3), com raízes complexas, a condição parece ser mais especializada; ela estipula somente que a parte real (h) das raízes complexas ($h \pm vi$) seja negativa. Contudo, é possível unificar todos os três casos e consolidar as condições aparentemente diferentes em uma condição única, aplicável de modo geral. Interprete qualquer raiz real r como uma raiz complexa cuja parte imaginária é zero ($v = 0$). Então, a condição “a parte real de toda raiz característica é negativa” torna-se claramente aplicável a todos os três casos e surge como a única condição de que precisamos.

EXERCÍCIO 16.3

Encontre y_p e a y_c , a solução geral e a solução definida de cada uma das seguintes:

1. $y''(t) - 4y'(t) + 8y = 0$; $y(0) = 3$, $y'(0) = 7$
2. $y''(t) + 4y'(t) + 8y = 2$; $y(0) = 2\frac{1}{4}$, $y'(0) = 4$
3. $y''(t) + 3y'(t) + 4y = 12$; $y(0) = 2$, $y'(0) = 2$
4. $y''(t) - 2y'(t) + 10y = 5$; $y(0) = 6$, $y'(0) = 8\frac{1}{2}$
5. $y''(t) + 9y = 3$; $y(0) = 1$, $y'(0) = 3$
6. $2y''(t) - 12y'(t) + 20y = 40$; $y(0) = 4$, $y'(0) = 5$
7. Quais das equações diferenciais nos Problemas 1 a 6 dão trajetórias temporais com (a) flutuação amortecida; (b) flutuação uniforme; (c) flutuação explosiva?

16.4 Um modelo de mercado com expectativas de preço

Na formulação anterior do modelo dinâmico de mercado, Q_d e Q_s são ambas tomadas como funções apenas do preço corrente P . Mas, às vezes, compradores e vendedores podem basear seu comportamento de mercado não somente no preço corrente, mas também na *tendência* de preço dominante na ocasião, pois a tendência de preço provavelmente despertará neles certas *expectativas* em relação ao nível de preço no futuro, e essas expectativas podem, por sua vez, influenciar suas decisões de demanda e oferta.

Tendência de preço e expectativas de preço

No contexto do tempo contínuo, encontramos a informação da tendência de preço primariamente nas duas derivadas dP/dt (se o preço está subindo) e d^2P/dt^2 (se está subindo a uma taxa crescente). Para levar em conta a tendência de preço, agora vamos incluir essas derivadas como argumentos adicionais nas funções demanda e oferta:

$$Q_d = D[P(t), P'(t), P''(t)]$$

$$Q_s = S[P(t), P'(t), P''(t)]$$

Se nos limitarmos à versão linear dessas funções e simplificarmos a notação para as variáveis independentes para P , P' e P'' , podemos escrever

$$\begin{aligned} Q_d &= \alpha - \beta P + mP' + nP'' & (\alpha, \beta > 0) \\ Q_s &= -\gamma + \delta P + uP' + wP'' & (\gamma, \delta > 0) \end{aligned} \quad (16.26)$$

onde os parâmetros α , β , γ , e δ são meras importações dos modelos de mercado anteriores, mas m , n , u e w são novos.

Os quatro novos parâmetros, a cujos sinais não foram impostas restrições, incorporam as expectativas de preço de compradores e vendedores. Se $m > 0$, por exemplo, um preço em elevação fará com que Q_d aumente. Isso sugeriria que os compradores esperam que o preço que está subindo *continue* a subir e, por consequência, preferem aumentar suas compras agora, quando o preço ainda está relativamente baixo. O sinal oposto para m , por outro lado, significaria a expectativa de uma reversão imediata da tendência de preço, portanto os compradores prefeririam reduzir as compras correntes e esperar que um preço mais baixo se materializasse mais tarde. A inclusão do parâmetro n faz com que o comportamento dos compradores dependa também da taxa de variação de dP/dt . Assim, os novos parâmetros m e n injetam um elemento substancial de especulação de preços no modelo. Os parâmetros u e w acarretam uma implicação similar no lado dos vendedores desse quadro.

Um modelo simplificado

Por simplicidade, vamos admitir que somente a função demanda contenha expectativas de preço. Especificamente, sejam m e n diferentes de zero, mas deixemos que $u = w = 0$ em (16.26). Vamos supor, ainda, que o mercado é compensado a cada instante. Então podemos igualar as funções demanda e oferta para obter (após normalização) a equação diferencial

$$P'' + \frac{m}{n}P' - \frac{\beta + \delta}{n}P = -\frac{\alpha + \gamma}{n} \quad (16.27)$$

Essa equação está na forma de (16.2) com as seguintes substituições:

$$y = P \quad a_1 = \frac{m}{n} \quad a_2 = -\frac{\beta + \delta}{n} \quad b = -\frac{\alpha + \gamma}{n}$$

Uma vez que esse padrão de variação de P envolve a derivada segunda P'' bem como a derivada primeira P' , o presente modelo é certamente distinto do modelo dinâmico de mercado apresentado na Seção 15.2.

Todavia, note que o presente modelo é diferente do modelo anterior de um outro modo. Na Seção 15.2, está presente um mecanismo de ajuste dinâmico, $dP/dt = j(Q_d - Q_s)$. Uma vez que essa equação implica que $dP/dt = 0$ se, e somente se, $Q_d = Q_s$, o sentido intertemporal e o sentido de compensação de mercado do equilíbrio coincidem naquele modelo. O presente modelo, ao contrário, admite compensação de mercado a cada instante de tempo. Assim, cada preço atingido no mercado é um preço de equilíbrio no sentido de compensação de mercado, embora possa não se qualificar como o preço de equilíbrio intertemporal. Em outras palavras, os dois sentidos de equilíbrio agora são díspares. Note, também, que o mecanismo de ajuste $dP/dt = j(Q_d - Q_s)$, contendo uma derivada, é o que faz do modelo de mercado anterior um modelo dinâmico. No presente modelo, sem nenhum mecanismo de ajuste, a natureza dinâmica do modelo emana, por sua vez, dos termos de expectativa mP' e nP'' .

A trajetória temporal do preço

O preço de equilíbrio intertemporal desse modelo – a solução particular P_p (anteriormente y_p) – é encontrado com facilidade usando (16.3). Ele é

$$P_p = \frac{b}{a_2} = \frac{\alpha + \gamma}{\beta + \delta}$$

Como essa é uma constante (positiva), ela representa um equilíbrio constante.

Quanto à função complementar P_c (anteriormente y_c), há três casos possíveis.

Caso 1 (raízes reais distintas)

$$\left(\frac{m}{n}\right)^2 > -4\left(\frac{\beta + \delta}{n}\right)$$

A função complementar deste caso é, por (16.7),

$$P_c = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$

onde

$$r_1, r_2 = \frac{1}{2} \left[-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta + \delta}{n}\right)} \right] \quad (16.28)$$

De acordo com isso, a solução geral é

$$P(t) = P_c + P_p = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \frac{\alpha + \gamma}{\beta + \delta} \quad (16.29)$$

Caso 2 (raízes reais duplas)

$$\left(\frac{m}{n}\right)^2 = -4\left(\frac{\beta + \delta}{n}\right)$$

Neste caso, as raízes características assumem o valor único

$$r = -\frac{m}{2n}$$

assim, por (16.9), a solução geral pode ser escrita como

$$P(t) = A_3 e^{-mt/2n} + A_4 t e^{-mt/2n} + \frac{\alpha + \gamma}{\beta + \delta} \quad (16.29')$$

Caso 3 (raízes complexas)

$$\left(\frac{m}{n}\right)^2 < -4\left(\frac{\beta + \delta}{n}\right)$$

Neste terceiro e último caso, as raízes características são o par de números complexos conjugados

$$r_1, r_2 = h \pm vi$$

onde

$$h = -\frac{m}{2n} \quad \text{e} \quad v = \frac{1}{2} \sqrt{-4\left(\frac{\beta + \delta}{n}\right) - \left(\frac{m}{n}\right)^2}$$

Por conseguinte, por (16.24'), temos a solução geral

$$P(t) = e^{-mt/2n} (A_5 \cos vt + A_6 \sin vt) + \frac{\alpha + \gamma}{\beta + \delta} \quad (16.29'')$$

Podemos deduzir algumas conclusões gerais desses resultados. Primeiro, se $n > 0$, então $-4(\beta + \delta)/n$ deve ser negativo e, portanto, menor que $(m/n)^2$. Por conseguinte, os Casos 2 e 3 podem ser descartados imediatamente. Além disso, com n positivo (como o são β e δ), a expressão sob o sinal de raiz quadrada em (16.28) é necessariamente maior que $(m/n)^2$ e, assim, a raiz quadrada deve ser maior que $|m/n|$. O sinal $\pm$ em (16.28) então produziria uma raiz positiva (r_1) e uma raiz negativa (r_2). Por consequência, o equilíbrio intertemporal é dinamicamente instável, a menos que o valor definido da constante A_1 seja, por acaso, zero em (16.29).

Em segundo lugar, se $n < 0$, então todos os três casos tornam-se viáveis. Sob o Caso 1, podemos ter certeza de que ambas as raízes serão negativas se m for negativo. (Por quê?) O interessante é que a raiz repetida do Caso 2 também será negativa se m for negativo. Além disso, visto que h , a parte real das raízes complexas no Caso 3, assume o mesmo valor que as raízes repetidas r no Caso 2, a negatividade de m também garantirá que h seja negativo. Em suma, para todos os três casos, a estabilidade dinâmica de equilíbrio é garantida quando os parâmetros m e n forem ambos negativos.

EXEMPLO 1

Sejam as funções demanda e oferta

$$Q_d = 42 - 4P - 4P' + P''$$

$$Q_s = -6 + 8P$$

com condições iniciais $P(0) = 6$ e $P'(0) = 4$. Supondo uma compensação de mercado em cada instante, encontre a trajetória temporal $P(t)$.

Neste exemplo, os valores dos parâmetros são

$$\alpha = 42 \quad \beta = 4 \quad \gamma = 6 \quad \delta = 8 \quad m = -4 \quad n = 1$$

Visto que n é positivo, nossa discussão anterior sugere que somente o Caso 1 pode surgir, e que as duas raízes (reais) r_1 e r_2 assumirão sinais opostos. A substituição dos valores dos parâmetros em (16.28) realmente confirma isso, pois

$$r_1, r_2 = \frac{1}{2}(4 \pm \sqrt{16 + 48}) = \frac{1}{2}(4 \pm 8) = 6, -2$$

Então, por (16.29), a solução geral é

$$P(t) = A_1 e^{6t} + A_2 e^{-2t} + 4$$

Além do mais, levando em conta as condições iniciais, constatamos que $A_1 = A_2 = 1$, portanto, a solução definida é

$$P(t) = e^{6t} + e^{-2t} + 4$$

Em vista da raiz positiva $r_1 = 6$, o equilíbrio intertemporal ($P_p = 4$) é dinamicamente instável.

A solução precedente é encontrada por meio da utilização das fórmulas (16.28) e (16.29). Alternativamente, podemos primeiro igualar as funções demanda e oferta dadas para obter a equação diferencial

$$P'' - 4P' - 12P = -48$$

e então resolver essa equação como um caso específico de (16.2).

EXEMPLO 2

Dadas as funções demanda e oferta

$$Q_d = 40 - 2P - 2P' - P''$$

$$Q_s = -5 + 3P$$

com $P(0) = 12$ e $P'(0) = 1$, encontre $P(t)$ supondo que o mercado é sempre compensado.

Aqui, os parâmetros m e n são ambos negativos. Portanto, de acordo com nossa discussão geral anterior, o equilíbrio intertemporal deve ser dinamicamente estável. Para encontrar a solução específica, podemos primeiro igualar Q_d e Q_s para obter a equação diferencial (após multiplicar tudo por -1)

$$P'' + 2P' + 5P = 45$$

O equilíbrio intertemporal é dado pela solução particular

$$P_p = \frac{45}{5} = 9$$

Pela equação característica da equação diferencial,

$$r^2 + 2r + 5 = 0$$

verificamos que as raízes são complexas:

$$r_1, r_2 = \frac{1}{2}(-2 \pm \sqrt{4-20}) = \frac{1}{2}(-2 \pm 4i) = -1 \pm 2i$$

Isso significa que $h = -1$ e $v = 2$, portanto a solução geral é

$$P(t) = e^{-t}(A_5 \cos 2t + A_6 \sin 2t) + 9$$

Para definir as constantes arbitrárias A_5 e A_6 , estabelecemos $t = 0$ na solução geral, para obter

$$P(0) = e^0(A_5 \cos 0 + A_6 \sin 0) + 9 = A_5 + 9 \quad [\cos 0 = 1; \sin 0 = 0]$$

Além disso, diferenciando a solução geral e então estabelecendo $t = 0$, constatamos que

$$P'(t) = -e^{-t}(A_5 \cos 2t + A_6 \sin 2t) + e^{-t}(-2A_5 \sin 2t + 2A_6 \cos 2t)$$

[regra do produto e regra da cadeia]

$$\begin{aligned} \text{e} \quad P'(0) &= -e^0(A_5 \cos 0 + A_6 \sin 0) + e^0(-2A_5 \sin 0 + 2A_6 \cos 0) \\ &= -(A_5 + 0) + (0 + 2A_6) = -A_5 + 2A_6 \end{aligned}$$

Assim, em virtude das condições iniciais $P(0) = 12$ e $P'(0) = 1$, temos $A_5 = 3$ e $A_6 = 2$. Por consequência, a solução definida é

$$P(t) = e^{-t}(3 \cos 2t + 2 \sin 2t) + 9$$

Essa trajetória temporal é obviamente uma trajetória de flutuação periódica; o período é $2\pi/v = \pi$. Isto é, há um ciclo completo toda vez que t aumentar de $\pi = 3,14159\dots$. Em vista do termo multiplicativo e^{-t} , a flutuação é amortecida. A trajetória temporal, que começa no preço inicial $P(0) = 12$, converge para o preço de equilíbrio intertemporal $P_p = 9$ de um modo cíclico.

EXERCÍCIO 16.4

- Sejam os parâmetros m , n , u e w em (16.26) todos diferentes de zero.
 - Supondo uma compensação de mercado em cada instante de tempo, escreva a nova equação diferencial do modelo.
 - Encontre o preço de equilíbrio intertemporal.
 - Sob quais circunstâncias a flutuação periódica pode ser excluída?
- Sejam as funções demanda e oferta como em (16.26), mas com $u = w = 0$ como na discussão no texto da discussão.
 - Se o mercado não for sempre compensado, mas se ajustar de acordo com

$$\frac{dP}{dt} = j(Q_d - Q_s) \quad (j > 0)$$

escreva a nova equação diferencial apropriada.

- Encontre o preço de equilíbrio intertemporal $\bar{P}$ e o preço de equilíbrio de compensação de mercado P^* .
 - Enuncie a condição para se ter uma trajetória de preço flutuante. Pode ocorrer flutuação se $n > 0$?
- Sejam a demanda e a oferta

$$Q_d = 9 - P + P' + 3P'' \quad Q_s = -1 + 4P - P' + 5P''$$

com $P(0) = 4$ e $P'(0) = 4$.

 - Encontre a trajetória de preço supondo uma compensação de mercado em cada instante.
 - A trajetória temporal é convergente? Com flutuação?

16.5 A interação entre inflação e desemprego

Nesta seção, ilustraremos a utilização de uma equação diferencial de segunda ordem com um macromodelo que trata do problema da inflação e do desemprego.

A relação de Phillips

Um dos conceitos mais amplamente utilizados na análise moderna do problema da inflação e do desemprego é a relação de Phillips.[†] Em sua formulação original, essa relação retrata uma relação negativa obtida empiricamente entre a taxa de crescimento da renda salarial e a taxa de desemprego:

$$w = f(U) \quad [f'(U) < 0] \quad (16.30)$$

onde a letra minúscula w denota a taxa de crescimento da renda salarial W (isto é, $w \equiv \dot{W}/W$) e U é a taxa de desemprego. Assim, essa relação é pertinente exclusivamente ao mercado de trabalho. A utilização posterior, contudo, adaptou a relação de Phillips para uma função que liga a *taxa de inflação* (em vez de w) à taxa de desemprego. Essa adaptação pode ser justificada argumentando que a remarcação de preços é amplamente utilizada, de modo que um w positivo, que reflete o custo crescente da renda salarial, necessariamente acarretaria implicações inflacionárias. E isso faz da taxa de inflação, assim como w , uma função de U . Contudo, a pressão inflacionária de um w positivo pode ser compensada por um aumento na produtividade do trabalho, admitida como exógena e denotada aqui por T . Especificamente, o efeito inflacionário pode se materializar somente até o ponto em que a renda salarial crescer mais rapidamente do que a produtividade. Denotando a taxa de inflação – isto é, a taxa de crescimento do nível de preço P – pela letra minúscula p , ($p \equiv \dot{P}/P$), podemos escrever

$$p = w - T \quad (16.31)$$

Combinando (16.30) e (16.31), e adotando a versão linear da função $f(U)$, obtemos, então, uma relação de Phillips adaptada

$$p = \alpha - T - \beta U \quad (\alpha, \beta > 0) \quad (16.32)$$

A relação de Phillips com expectativas aumentadas

De uns tempos para cá, os economistas têm preferido a versão da relação de Phillips com *expectativas aumentadas*.

$$w = f(U) + g\pi \quad (0 < g \leq 1) \quad (16.30')$$

onde π denota a taxa de inflação esperada. Então, a idéia subjacente a (16.30'), como proposta pelo professor Friedman, agraciado com um Prêmio Nobel,^{††} é que, se uma tendência inflacionária se mantiver em efeito por tempo suficiente, as pessoas estarão propensas a formar certas expectativas de inflação que então tentam incorporar às suas demandas de renda salarial. Assim, w deve ser uma função crescente de π . Transposta para (16.32), essa idéia resulta na equação

$$p = \alpha - T - \beta U + g\pi \quad (0 < g \leq 1) \quad (16.33)$$

[†] Phillips, A. W. "The Relationship Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957". *Economica*, p. 283–299, nov. 1958.

^{††} Friedman, Milton. "The Role of Monetary Policy". *American Economic Review*, p. 1–17, mar. 1968.

Com a introdução de uma nova variável para denotar a taxa de inflação esperada, torna-se necessário propor hipóteses sobre como as expectativas de inflação são especificamente formadas.[†] Aqui, adotamos a hipótese das *expectativas adaptativas*

$$\frac{d\pi}{dt} = j(p - \pi) \quad (0 < j \leq 1) \quad (16.34)$$

Note que, em vez de explicar a grandeza absoluta de π , essa equação descreve seu padrão de variação ao longo do tempo. Se acontecer de a taxa real de inflação p exceder a taxa esperada π , esta última, que agora demonstrou ser muito baixa, é revisada para cima ($d\pi/dt > 0$). Reciprocamente, se p ficar abaixo de π , então π é revisada na direção da redução. O formato de (16.34) lembra muito o mecanismo de ajuste $dP/dt = j(Q_d - Q_s)$ do modelo de mercado. Mas aqui, a força propulsora subjacente ao ajuste é a discrepância entre as taxas de inflação *real* e *esperada*, e não entre Q_d e Q_s .

O retorno da inflação para o desemprego

Pode-se considerar que (16.33) e (16.34) constituem um modelo completo. Entretanto, uma vez que há três variáveis em um sistema de duas equações, uma delas tem de ser considerada exógena. Se π e p forem consideradas endógenas, por exemplo, então U deve ser tratada como exógena. Uma alternativa mais satisfatória é introduzir uma terceira equação para explicar a variável U , de modo a enriquecer o modelo com características comportamentais. O mais significativo é que isso nos proporcionará uma oportunidade de levar em conta o efeito de retorno da inflação sobre o desemprego. A equação (16.33) nos mostra como U afeta p – em grande parte pelo lado da oferta da economia. Mas é certo que p , por sua vez, também pode afetar U . Por exemplo, a taxa de inflação pode influenciar as decisões de poupança de consumo do público e, por consequência, também a demanda agregada pela produção interna, e esta última, por sua vez, afetará a taxa de desemprego. A taxa de inflação pode até mesmo provocar uma diferença na efetividade da condução de políticas governamentais de gerenciamento de demanda. Dependendo da taxa de inflação, um dado nível de dispêndio de moeda (política fiscal) poderia ser traduzido em níveis variáveis de dispêndio real e, de modo semelhante, uma dada taxa de expansão da moeda nominal (política monetária) poderia significar taxas variáveis de expansão da moeda real. E essas, por sua vez, implicariam efeitos diferentes sobre a produção e o desemprego.

Por simplicidade, levaremos em consideração apenas o retorno por meio da condução da política monetária. Denotando o equilíbrio da moeda nominal por M e sua taxa de crescimento por $m \equiv \dot{M}/M$, vamos postular que^{††}

$$\frac{dU}{dt} = -k(m - p) \quad (k > 0) \quad (16.35)$$

Aplicando (10.25) de trás para a frente, vemos que a expressão $(m - p)$ representa a taxa de crescimento da moeda *real*:

$$m - p = \frac{\dot{M}}{M} - \frac{\dot{P}}{P} = r_M - r_P = r_{(M/P)}$$

Assim, (16.35) estipula que dU/dt está negativamente relacionada à taxa de crescimento do equilíbrio de moeda real. Considerando que a variável p agora entra na determinação de dU/dt , o modelo agora contém um retorno da inflação sobre o desemprego.

[†] Isso é o contrário da Seção 16.4, em que as expectativas de preço foram discutidas sem a introdução de uma nova variável para representar o preço esperado. O resultado é que as premissas referentes à formação de expectativas estavam apenas implicitamente embutidas nos parâmetros m , n , u e w em (16.26).

^{††} Em uma discussão anterior, denotamos a oferta de moeda por M_s , para distingui-la da demanda por moeda M_d . Aqui, podemos simplesmente usar a letra M , sem índice, já que não há perigo de confusão.

A trajetória temporal do π

Juntas, as expressões (16.33), (16.34) e (16.35) constituem um modelo fechado nas três variáveis π , p e U . Eliminando duas das três variáveis, contudo, podemos condensar o modelo em uma única equação diferencial de uma única variável. Suponha que deixemos que a única variável seja π . Então, primeiro substituímos (16.33) em (16.34) para obter

$$\frac{d\pi}{dt} = j(\alpha - T - \beta U) - j(1 - g)\pi \quad (16.36)$$

Se essa equação contivesse a expressão dU/dt em vez de U , poderíamos ter substituído (16.35) em (16.36) diretamente. Mas, com a configuração de (16.36), em primeiro lugar devemos criar deliberadamente um termo dU/dt pela diferenciação de (16.36) em relação a t , com o resultado

$$\frac{d^2\pi}{dt^2} = -j\beta \frac{dU}{dt} - j(1 - g) \frac{d\pi}{dt} \quad (16.37)$$

A substituição de (16.35) nessa expressão então resulta em

$$\frac{d^2\pi}{dt^2} = j\beta km - j\beta kp - j(1 - g) \frac{d\pi}{dt} \quad (16.37')$$

Ainda há uma variável p a ser eliminada. Para fazer isso, notamos que (16.34) implica

$$p = \frac{1}{j} \frac{d\pi}{dt} + \pi \quad (16.38)$$

Usando esse resultado em (16.37') e simplificando, finalmente obtemos a equação diferencial desejada na variável π apenas:

$$\frac{d^2\pi}{dt^2} + \underbrace{[\beta k + j(1 - g)]}_{a_1} \frac{d\pi}{dt} + \underbrace{(j\beta k)}_{a_2} \pi = \underbrace{j\beta km}_b \quad (16.37'')$$

A solução particular dessa equação é simplesmente

$$\pi_p = \frac{b}{a_2} = m$$

Assim, nesse modelo, o valor do equilíbrio intertemporal da taxa de inflação esperada depende exclusivamente da taxa de crescimento da moeda nominal.

Para a função complementar, as duas raízes são, como antes,

$$r_1, r_2 = \frac{1}{2} \left(-a_1 \pm \sqrt{a_1^2 - 4a_2} \right) \quad (16.39)$$

onde, como pode ser notado por (16.37''), a_1 e a_2 são ambos positivos. Não seria possível determinar *a priori* se a_2 seria maior, igual ou menor que $4a_2$. Assim, todos os três casos de raízes características podem surgir – raízes reais distintas, raízes reais repetidas ou raízes complexas. Entretanto, seja qual for o caso que se apresente, o equilíbrio intertemporal revelará ser dinamicamente estável no presente modelo. Isso pode ser explicado da seguinte maneira: suponha, em primeiro lugar, que prevaleça o Caso 1, com $a_1^2 > 4a_2$. Então, a raiz quadrada em (16.39) dá um número real. Visto que a_2 é positivo, $\sqrt{a_1^2 - 4a_2}$ é necessariamente menor que $\sqrt{a_1^2} = a_1$. Deduz-se que r_1 é negativa, assim como r_2 , o que implica um equilíbrio dinamicamente estável. E se $a_1^2 = 4a_2$ (Caso 2)? Nesse caso, a raiz quadrada é zero, de modo que $r_1 = r_2 = -a_1/2 < 0$. E a nega-

tividade das raízes repetidas novamente implica estabilidade dinâmica. Finalmente, para o Caso 3, a parte real das raízes complexas é $h = -a_1/2$. Uma vez que esta tem o mesmo valor das raízes repetidas no Caso 2, aplica-se a conclusão idêntica à que chegamos em relação à estabilidade dinâmica.

Embora tenhamos estudado apenas a trajetória temporal de π , o modelo certamente pode render informações sobre as outras variáveis também. Para encontrar, por exemplo, a trajetória temporal da variável U , podemos começar *ou* condensando o modelo em uma equação diferencial em U e não em π (veja Exercício 16.5–2) *ou* deduzir a trajetória de U pela trajetória de π que já encontramos (veja Exemplo 1).

EXEMPLO 1 Façamos com que as três equações do modelo tomem as formas específicas

$$p = \frac{1}{6} - 3U + \pi \quad (16.40)$$

$$\frac{d\pi}{dt} = \frac{3}{4}(p - \pi) \quad (16.41)$$

$$\frac{dU}{dt} = -\frac{1}{2}(m - p) \quad (16.42)$$

Então temos os valores de parâmetros $\beta = 3$, $h = 1$, $j = \frac{3}{4}$ e $k = \frac{1}{2}$; assim, com referência a (16.37''), temos

$$a_1 = \beta k + j(1 - g) = \frac{3}{2} \quad a_2 = j\beta k = \frac{9}{8} \quad \text{e} \quad b = j\beta km = \frac{9}{8}m$$

A solução particular é $b/a_2 = m$. Com $a_1^2 < 4a_2$, as raízes características são complexas:

$$r_1, r_2 = \frac{1}{2} \left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}} \right) = \frac{1}{2} \left(-\frac{3}{2} \pm \frac{3}{2}i \right) = -\frac{3}{4} \pm \frac{3}{4}i$$

Isto é, $h = -\frac{3}{4}$ e $v = \frac{3}{4}$. Por consequência, a solução geral para a taxa de inflação esperada é

$$\pi(t) = e^{-3t/4} \left(A_5 \cos \frac{3}{4}t + A_6 \sin \frac{3}{4}t \right) + m \quad (16.43)$$

que representa uma trajetória temporal com flutuação amortecida ao redor do valor de equilíbrio m .

Dessa expressão também podemos deduzir as trajetórias temporais para as variáveis p e U . De acordo com (16.41), p pode ser expresso em termos de π e $d\pi/dt$ pela equação

$$p = \frac{4}{3} \frac{d\pi}{dt} + \pi$$

A trajetória de π na solução geral (16.43) implica a derivada

$$\begin{aligned} \frac{d\pi}{dt} &= -\frac{3}{4}e^{-3t/4} \left(A_5 \cos \frac{3}{4}t + A_6 \sin \frac{3}{4}t \right) \\ &\quad + e^{-3t/4} \left(-\frac{3}{4}A_5 \sin \frac{3}{4}t + \frac{3}{4}A_6 \cos \frac{3}{4}t \right) \quad [\text{regra do produto e regra da cadeia}] \end{aligned}$$

Usando a solução (16.43) e sua derivada, temos, então,

$$p(t) = e^{-3t/4} \left(A_6 \cos \frac{3}{4}t - A_5 \sin \frac{3}{4}t \right) + m \quad (16.44)$$

Assim como a taxa de inflação *esperada* π , a taxa de inflação *real* p também tem uma trajetória temporal flutuante que converge para o valor de equilíbrio m .

Quanto à variável U , (16.40) nos informa que ela pode ser expressa em termos de π e p como se segue:

$$U = \frac{1}{3}(\pi - p) + \frac{1}{18}$$

Por conseguinte, em virtude das soluções (16.43) e (16.44), podemos escrever a trajetória temporal da taxa de desemprego como

$$U(t) = \frac{1}{3}e^{-3t/4} \left[(A_5 - A_6) \cos \frac{3}{4}t + (A_5 + A_6) \sin \frac{3}{4}t \right] + \frac{1}{18} \quad (16.45)$$

Mais uma vez, essa é uma trajetória amortecida, com $\frac{1}{18}$ como $\bar{U}$, o valor de equilíbrio intertemporal dinamicamente estável de U .

Como os valores de equilíbrio intertemporal de π e p são ambos iguais ao parâmetro de política monetária m , o valor de m – a taxa de crescimento da moeda nominal – proporciona o eixo ao redor do qual as trajetórias temporais de π e p flutuam. Se ocorrer uma mudança em m , um novo valor de equilíbrio de π e p substituirá imediatamente o antigo e, sejam quais forem os valores que as variáveis π e p venham a assumir no momento da variação da política monetária, eles se tornarão os valores iniciais a partir dos quais emanam as novas trajetórias π e p .

Por outro lado, o valor do equilíbrio intertemporal $\bar{U}$ não depende de m . Segundo (16.45), U converge para a constante $\frac{1}{18}$ independentemente da taxa de crescimento da moeda nominal e, por conseguinte, independentemente da taxa de equilíbrio da inflação. Esse valor de equilíbrio constante de U é denominado a *taxa natural de desemprego*. O fato de a taxa natural de desemprego ser consistente com qualquer taxa de equilíbrio da inflação pode ser representado no espaço Up por uma reta vertical paralela ao eixo p . Essa reta vertical que relaciona os valores de equilíbrio de U e p entre si é conhecida como a *curva de Phillips de longo prazo*. A forma vertical dessa curva, entretanto, é contingente de um valor especial de parâmetro admitido neste exemplo. Quando o valor é alterado, como no Exercício 16.5–4, a curva de Phillips de longo prazo pode não ser mais vertical.

EXERCÍCIO 16.5

1. No modelo inflação-desemprego, retenha (16.33) e (16.34) mas esqueça (16.35) e deixe que U seja exógena, desta vez.
 - (a) Que tipo de equação diferencial surgirá agora?
 - (b) Quantas raízes características você consegue obter? Agora é possível ter flutuação periódica na função complementar?
2. Na discussão no texto, condensamos o modelo inflação-desemprego em uma equação diferencial na variável π . Mostre que, alternativamente, o modelo pode ser condensado em uma equação diferencial de segunda ordem na variável U , com os mesmos coeficientes a_1 e a_2 que aparecem em (16.37''), mas um termo constante diferente $b = kj[\alpha - T - (1 - g)m]$.
3. Substitua a hipótese de expectativas adaptativas (16.34) pela denominada hipótese da previsão perfeita $\pi = p$, mas mantenha (16.33) e (16.35).
 - (a) Obtenha uma equação diferencial na variável p .
 - (b) Obtenha uma equação diferencial na variável U .
 - (c) Como essas equações diferem fundamentalmente daquela que obtivemos sob a hipótese de expectativas adaptativas?
 - (d) Qual mudança na restrição de parâmetro é necessária agora para tornar a equação diferencial significativa?

4. No Exemplo 1, mantenha (16.41) e (16.42), mas substitua (16.40) por

$$p = \frac{1}{6} - 3U + \frac{1}{3}\pi$$

- (a) Encontre $p(t)$, $\pi(t)$ e $U(t)$.
 (b) As trajetórias temporais ainda estão flutuando? Ainda são convergentes?
 (c) Quais são $\bar{p}$ e $\bar{U}$, os valores de equilíbrio intertemporal de p e U ?
 (d) Ainda é verdade que $\bar{U}$ não é relacionado funcionalmente com $\bar{p}$? Se agora ligarmos esses dois valores de equilíbrio entre si em uma curva de Phillips de longo prazo, ainda podemos obter uma curva vertical? Assim, qual das premissas no Exemplo 1 é crucial se obter derivar uma curva de Phillips de longo prazo vertical?

16.6 Equações diferenciais com um termo variável

Nas equações diferenciais consideradas na Seção 16.1,

$$y''(t) + a_1 y'(t) + a_2 y = b$$

o termo b do lado direito é uma constante. E se, em vez de b , nós tivéssemos um *termo variável* na direita; isto é, alguma função de t tal como bt^2 , e^{bt} ou $b \sin t$? A resposta é que, então, teríamos de modificar nossa solução particular y_p . Felizmente, a função complementar não é afetada pela presença de um termo variável, porque y_c trata somente da equação homogênea associada, cujo lado direito é sempre zero.

Método de coeficientes indeterminados

Explicaremos um método para obter y_p , conhecido como o *método de coeficientes indeterminados*, que é aplicável a equações diferenciais de coeficiente constante e termo variável, contanto que o termo variável e suas sucessivas derivadas contenham, em conjunto, somente um número *finito* de tipos distintos de expressões (à parte as constantes multiplicativas). A explicação para esse método pode ser melhor executada com uma ilustração concreta.

EXEMPLO 1 Encontre a solução particular de

$$y''(t) + 5y'(t) + 3y = 6t^2 - t - 1 \quad (16.46)$$

Por definição, a solução particular é um valor de y que satisfaz a equação dada, isto é, um valor de y que fará com que o lado esquerdo seja exatamente igual ao lado direito independentemente do valor de t . Visto que o lado esquerdo contém a função $y(t)$ e as derivadas $y'(t)$ e $y''(t)$ – ao passo que o lado direito contém múltiplos das expressões t^2 , t e uma constante –, perguntamos: qual forma geral de função de $y(t)$, juntamente com suas derivadas primeira e segunda, nos darão os três tipos de expressões t^2 , t e uma constante? A resposta óbvia é uma função da forma $B_1 t^2 + B_2 t + B_3$ (onde B_i são coeficientes ainda a serem determinados) pois, se escrevermos a solução particular como

$$y(t) = B_1 t^2 + B_2 t + B_3$$

obtemos

$$y'(t) = 2B_1 t + B_2 \quad \text{e} \quad y''(t) = 2B_1 \quad (16.47)$$

e essas três equações são, de fato, compostas dos tipos de expressões citados. Substituindo essas expressões em (16.46) e agrupando os termos, obtemos

$$\text{Lado esquerdo} = (3B_1)t^2 + (10B_1 + 3B_2)t + (2B_1 + 5B_2 + 3B_3)$$

E, quando igualamos essa equação termo a termo com o lado direito, podemos determinar os coeficientes B_i como se segue:

$$\begin{cases} 3B_1 = 6 \\ 10B_1 + 3B_2 = -1 \\ 2B_1 + 5B_2 + 3B_3 = -1 \end{cases} \Rightarrow \begin{cases} B_1 = 2 \\ B_2 = -7 \\ B_3 = 10 \end{cases}$$

Assim, a solução particular desejada pode ser escrita como

$$y_p = 2t^2 - 7t + 10$$

Esse método pode funcionar somente quando o número de tipos de expressão for finito (Veja Exercício 16.6–1.) Em geral, quando esse pré-requisito é cumprido, podemos considerar que a solução particular tem a forma de uma combinação linear de todos os tipos de expressões contidos no termo variável dado, bem como em todas as suas derivadas. Observe, em particular, que uma expressão constante deve ser incluída na solução particular, se o termo variável original ou qualquer uma de suas derivadas sucessivas contiver um termo constante.

EXEMPLO 2

Como uma ilustração a mais, vamos encontrar a forma geral para a solução particular adequada para o termo variável $(b \sin t)$. Neste caso, uma diferenciação repetida resulta nas derivadas sucessivas $(b \cos t)$, $(-b \sin t)$, $(-b \cos t)$, $(b \sin t)$ etc., que envolvem somente dois tipos distintos de expressões. Portanto, podemos experimentar uma solução particular na forma $(B_1 \sin t + B_2 \cos t)$.

Uma modificação

Em certos casos, surge uma complicação na aplicação do método. Quando o coeficiente do termo y na equação diferencial dada for zero, tal como em

$$y''(t) + 5y'(t) = 6t^2 - t - 1$$

a forma experimental que utilizamos anteriormente para y_p , a saber, $B_1 t^2 + B_2 t + B_3$, não funcionará. A causa dessa falha é que, uma vez que o termo $y(t)$ está fora do quadro e visto que somente derivadas $y'(t)$ e $y''(t)$ do tipo mostrado em (16.47) serão substituídas no lado esquerdo, jamais aparecerá na esquerda um termo da forma $B_1 t^2$ para ser igualado ao termo $6t^2$ na direita. A saída para esse tipo de dificuldade é usar, em seu lugar, a solução experimental $t(B_1 t^2 + B_2 t + B_3)$; ou, se isso também falhar (por exemplo, dada a equação $y''(t) = 6t^2 - t - 1$), usar $t^2(B_1 t^2 + B_2 t + B_3)$, e assim por diante.

Na verdade, o mesmo truque pode ser empregado em uma outra circunstância difícil, como ilustrado no Exemplo 3.

EXEMPLO 2

Encontre a solução particular de

$$y''(t) + 3y'(t) - 4y = 2e^{-4t} \quad (16.48)$$

Aqui, o termo variável está na forma de e^{-4t} , mas todas as suas derivadas sucessivas (a saber, $-8e^{-4t}$, $32e^{-4t}$, $-128e^{-4t}$ etc.) também assumem a mesma forma. Se tentarmos a solução

$$y(t) = Be^{-4t} \quad [\text{com } y'(t) = -4Be^{-4t} \text{ e } y''(t) = 16Be^{-4t}]$$

e substituirmos essas expressões em (16.48), obtemos o inglório resultado

$$\text{Lado esquerdo} = (16 - 12 - 4)Be^{-4t} = 0 \quad (16.49)$$

que, obviamente, não pode ser igualado ao termo do lado direito $2e^{-4t}$.

O que faz com que isso aconteça é o fato de que o coeficiente exponencial no termo variável (-4) é, por acaso, igual a uma das raízes da equação característica de (16.48):

$$r^2 + 3r - 4 = 0 \quad (\text{raízes } r_1, r_2 = 1, -4)$$

A equação característica, como devemos recordar, é obtida por um processo de diferenciação;[†] mas a expressão $(16 - 12 - 4)$ em (16.49) é obtida pelo mesmo processo. Por conseguinte, não é nenhuma surpresa que $(16 - 12 - 4)$ seja uma mera versão específica de $(r^2 + 3r - 4)$ com r igualado a -4 . Visto que -4 , por acaso, é uma raiz característica, a expressão quadrática

$$r^2 + 3r - 4 = 16 - 12 - 4$$

deve ser, por necessidade, identicamente zero.

Para lidar com essa situação, vamos tentar, por sua vez, a solução

$$y(t) = Bte^{-4t}$$

com derivadas

$$y'(t) = (1 - 4t)Be^{-4t} \quad \text{e} \quad y''(t) = (-8 + 16t)Be^{-4t}$$

Substituindo essas expressões em (16.48), teremos agora: lado esquerdo $= -5Be^{-4t}$. Quando essa expressão é igualada ao lado direito, determinamos que o coeficiente é $B = -2/5$. Por conseguinte, a solução particular desejada de (16.48) pode ser escrita como

$$y_p = \frac{-2}{5}te^{-4t}$$

EXERCÍCIO 16.6

- Mostre que o método de coeficientes indeterminados não é aplicável à equação diferencial $y''(t) + ay'(t) + by = t^{-1}$.
- Encontre a solução particular de cada uma das seguintes equações pelo método dos coeficientes indeterminados:

(a) $y''(t) + 2y'(t) + y = t$	(c) $y''(t) + y'(t) + 2y = e^t$
(b) $y''(t) + 4y'(t) + y = 2t^2$	(d) $y''(t) + y'(t) + 3y = \sin t$

16.7 Equações diferenciais lineares de ordem mais alta

Os métodos de solução apresentados nas seções anteriores são estendidos sem muita dificuldade a uma equação diferencial linear de n -ésima ordem. Com coeficientes constantes e um termo constante, tal equação pode ser escrita, de forma geral, como

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y'(t) + a_n y = b \quad (16.50)$$

Encontrando a solução

Nesse caso de coeficientes constantes e termo constante, a presença de derivadas de ordem mais alta não afeta materialmente o método de achar a solução particular discutida anteriormente.

Se tentarmos o tipo de solução mais simples possível, $y = k$, podemos ver que todas as derivadas de $y'(t)$ a $y^{(n)}(t)$ serão zero; por conseguinte, (16.50) será reduzida a $a_n k = b$, e podemos escrever

[†] Veja a discussão que levou a (16.4").

$$y_p = k = \frac{b}{a_n} \quad (a_n \neq 0) \quad [\text{conforme (16.3)}]$$

Contudo, no caso de $a_n = 0$, devemos tentar uma solução da forma $y = kt$. Então, visto que $y'(t) = k$, todas as derivadas de ordem mais alta se anularão, (16.50) pode ser reduzida a $a_{n-1}k = b$, resultando, por conseguinte, na solução particular

$$y_p = kt = \frac{b}{a_{n-1}} t \quad (a_n = 0; a_{n-1} \neq 0) \quad [\text{conforme (16.3')}]$$

Se acontecer de $a_n = a_{n-1} = 0$, então esta última solução também falhará; em vez disso, deve ser tentada uma solução da forma $y = kt^2$. São óbvias as adaptações ulteriores desse procedimento.

Quanto à função complementar, a inclusão das derivadas de ordens mais altas na equação diferencial tem o efeito de elevar o grau da equação característica. A função complementar é definida como a solução geral da equação homogênea associada

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y'(t) + a_n y = 0 \quad (16.51)$$

Experimentando $y = Ae^{rt}$ ($\neq 0$) como uma solução e utilizando o conhecimento de que isso implica $y'(t) = r Ae^{rt}$, $y''(t) = r^2 Ae^{rt}$, ..., $y^{(n)}(t) = r^n Ae^{rt}$, podemos reescrever (16.51) como

$$Ae^{rt} (r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n) = 0$$

Essa equação é satisfeita por qualquer valor de r que satisfizer a seguinte equação característica (polinômio de grau n)

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0 \quad (16.51')$$

É claro que haverá n raízes para esse polinômio e cada uma delas deve ser incluída na solução geral de (16.51). Assim, nossa função complementar, de modo geral, deve ser da forma

$$y_c = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \cdots + A_n e^{r_n t} \quad \left(= \sum_{i=1}^n A_i e^{r_i t} \right)$$

Entretanto, como antes, devem ser feitas algumas modificações caso as n raízes não sejam todas reais e distintas. Em primeiro lugar, suponha que há raízes repetidas, digamos, $r_1 = r_2 = r_3$. Então, para evitar “redução”, devemos escrever os primeiros três termos das soluções como $A_1 e^{r_1 t} + A_2 t e^{r_1 t} + A_3 t^2 e^{r_1 t}$ [conforme (16.9)]. Caso tenhamos $r_4 = r_1$ também, o quarto termo deve ser alterado para $A_4 t^3 e^{r_1 t}$ etc.

Em segundo lugar, suponha que duas das raízes são complexas, digamos,

$$r_5, r_6 = h \pm vi$$

então, o quinto e o sexto termos da solução precedente devem ser combinados na seguinte expressão:

$$e^{ht} (A_5 \cos vt + A_6 \sin vt) \quad [\text{conforme (16.24')}]$$

Pelo mesmo critério, se forem achados dois pares *distintos* de raízes complexas, deve haver duas expressões trigonométricas (com um conjunto diferente de valores de h , v , e duas constantes arbitrárias para cada).[†] Mais uma outra possibilidade é, caso haja dois pares de raízes complexas *re-*

[†] É de interesse notar que, considerando-se que raízes complexas sempre vêm em pares conjugados, podemos ter certeza que teremos *ao menos uma* raiz real quando a equação diferencial for de ordem *ímpar*, isto é, quando n for um número ímpar.

petidas, então devemos usar e^{ht} como o termo multiplicativo para uma delas, mas usar te^{ht} para a outra. Além disso, mesmo que h e v tenham valores idênticos nas raízes complexas repetidas, agora deve ser atribuído um par diferente de constantes arbitrárias a cada uma.

Uma vez obtidas y_p e y_c , a solução geral da equação completa (16.50) é encontrada com facilidade. Como antes, ela é simplesmente a soma da função complementar com a solução particular: $y(t) = y_c + y_p$. Nessa solução geral, podemos contar um total de n constantes arbitrárias. Assim, para definir a solução, será exigido um total de n condições iniciais.

EXEMPLO 1 Ache a solução geral de

$$y^{(4)}(t) + 6y'''(t) + 14y''(t) + 16y'(t) + 8y = 24$$

A solução particular dessa equação de quarta ordem é simplesmente

$$y_p = \frac{24}{8} = 3$$

Sua equação característica é, por (16.51'),

$$r^4 + 6r^3 + 14r^2 + 16r + 8 = 0$$

que pode ser fatorada para a forma

$$(r + 2)(r + 2)(r^2 + 2r + 2) = 0$$

Das duas primeiras expressões paramétricas entre parênteses, podemos obter as raízes duplas $r_1 = r_2 = -2$, mas a última expressão (quadrática) dá como resultado o par de raízes complexas $r_3, r_4 = -1 \pm i$, com $h = -1$ e $v = 1$. Por consequência, a função complementar é

$$y_c = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t} (A_3 \cos t + A_4 \sin t)$$

e a solução geral é

$$y(t) = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t} (A_3 \cos t + A_4 \sin t) + 3$$

É claro que as quatro constantes A_1, A_2, A_3 e A_4 podem ser definidas se tivermos quatro condições iniciais.

Note que todas as raízes características neste exemplo são ou reais e negativas ou complexas com uma parte real negativa. Portanto, a trajetória temporal deve ser convergente e o equilíbrio intertemporal é dinamicamente estável.

Convergência e o teorema de Routh

A solução de uma equação característica de alto grau nem sempre é uma tarefa fácil. Por essa razão, seria uma tremenda ajuda se pudéssemos achar um modo de averiguar a convergência ou a divergência de uma trajetória temporal sem ter de resolver a equação e achar suas raízes características. Felizmente existe tal método, que pode fornecer uma análise qualitativa (se bem que não-gráfica) de uma equação diferencial.

Esse método é o *teorema de Routh*,[†] cujo enunciado é:

As partes reais de todas as raízes da equação polinomial de n -ésimo grau

$$a_0 r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$

[†] Se quiser consultar uma discussão desse teorema e um esboço de sua demonstração, consulte Samuelson, Paul A. *Foundations of Economic Analysis*. Harvard University Press, 1947, p. 429–435, e as referências ali citadas.

são negativas se, e somente se, os primeiros n determinantes da seguinte seqüência de determinantes

$$|a_1|; \quad \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}; \quad \dots$$

forem todos positivos.

Ao aplicar esse teorema, devemos lembrar que $|a_1| \equiv a_1$. Além disso, é preciso ficar entendido que devemos considerar $a_m = 0$ para todo $m > n$. Por exemplo, dada uma equação polinomial de terceiro grau ($n = 3$), precisamos examinar os sinais dos primeiros três determinantes listados no teorema de Routh; com essa finalidade, devemos estabelecer $a_4 = a_5 = 0$.

A relevância desse teorema para o problema da convergência deve ficar evidente por si só quando nos lembramos que, para a trajetória temporal $y(t)$ convergir independentemente de quais sejam as condições iniciais, todas as raízes características da equação diferencial devem ter partes reais negativas. Uma vez que a equação característica (16.51') é uma equação polinomial de n -ésimo grau, com $a_0 = 1$, o teorema de Routh pode ser um auxílio direto no teste de convergência. De fato, notamos que os coeficientes da equação característica (16.51') são totalmente idênticos aos da equação diferencial dada (16.51), portanto, é perfeitamente aceitável substituir os coeficientes de (16.51) diretamente na seqüência de determinantes mostrada no teorema de Routh para o teste, contanto que sempre tomemos $a_0 = 1$. Considerando que a condição citada no teorema é dada em termos de "se e somente se", é óbvio que ela constitui uma condição necessária e suficiente.

EXEMPLO 2 Teste, pelo teorema de Routh, se a equação diferencial do Exemplo 1 tem uma trajetória temporal convergente. Essa equação é de quarta ordem, portanto $n = 4$. Os coeficientes são $a_0 = 1$, $a_1 = 6$, $a_2 = 14$, $a_3 = 16$, $a_4 = 8$ e $a_5 = a_6 = a_7 = 0$. Substituindo esses coeficientes nos quatro primeiros determinantes, constatamos que seus valores são 6, 68, 800 e 6.400, respectivamente. Como todos eles são positivos, podemos concluir que a trajetória temporal é convergente.

EXERCÍCIO 16.7

- Encontre a solução particular de cada uma das seguintes:
 - $y'''(t) + 2y''(t) + y'(t) + 2y = 8$
 - $y'''(t) + y''(t) + 3y'(t) = 1$
 - $3y'''(t) + 9y''(t) = 1$
 - $y^{(4)}(t) + y''(t) = 4$
- Encontre y_p e y_c (e, por conseguinte, a solução geral) de:
 - $y'''(t) - 2y''(t) - y'(t) + 2y = 4$
[Sugestão: $r^3 - 2r^2 - r + 2 = (r - 1)(r + 1)(r - 2)$]
 - $y'''(t) + 7y''(t) + 15y'(t) + 9y = 0$
[Sugestão: $r^3 + 7r^2 + 15r + 9 = (r + 1)(r^2 + 6r + 9)$]
 - $y'''(t) + 6y''(t) + 10y'(t) + 8y = 8$
[Sugestão: $r^3 + 6r^2 + 10r + 8 = (r + 4)(r^2 + 2r + 2)$]
- Com base nos sinais das raízes características obtidas no Problema 2, analise a estabilidade de dinâmica de equilíbrio. Então verifique sua resposta pelo teorema de Routh.

4. Sem encontrar as raízes características das seguintes equações diferenciais, determine se elas darão origem a trajetórias temporais convergentes:
 - (a) $y''''(t) - 10y'''(t) + 27y''(t) - 18y' = 3$
 - (b) $y''''(t) + 11y'''(t) + 34y''(t) + 24y' = 5$
 - (c) $y''''(t) + 4y'''(t) + 5y''(t) - 2y' = -2$
5. Deduza, pelo teorema de Routh, que, para a equação diferencial linear de segunda ordem $y''(t) + a_1 y'(t) + a_2 y = b$, a trajetória de solução será convergente independentemente das condições iniciais se, e somente se, ambos os coeficientes a_1 e a_2 forem positivos.

CAPÍTULO 17

Tempo discreto: equações de diferenças de primeira ordem

No contexto do tempo contínuo, o padrão de variação de uma variável y é incorporado nas derivadas $y'(t)$, $y''(t)$ etc. A variação de tempo envolvida está ocorrendo continuamente. Por outro lado, quando o tempo é considerado uma variável *discreta*, de modo que a variável t só pode assumir valores inteiros, é óbvio que o conceito de derivada já não será mais apropriado. Então, como veremos, o padrão de variação da variável y deve ser descrito pelo que são denominadas diferenças de $y(t)$, e não por derivadas ou diferenciais. De acordo com isso, as técnicas de equações diferenciais darão lugar às técnicas de *equações de diferenças*.

Quando estamos tratando com tempo discreto, o valor da variável y mudará somente quando a variável t mudar de um valor inteiro para o seguinte, tal como de $t = 1$ para $t = 2$. No ínterim, supõe-se que nada acontece com y . Sob essa perspectiva, fica mais conveniente interpretar que os valores de t referem-se a *períodos* – e não a *instantes* – de tempo, e considerar que $t = 1$ denota período 1 e $t = 2$ denota período 2 e assim por diante. Então, podemos considerar simplesmente que y tem um único valor em cada período de tempo. Em vista dessa interpretação, a versão de tempo discreto da dinâmica econômica é frequentemente denominada *análise de período*. Devemos deixar bem claro, entretanto, que, aqui, “período” não está sendo usado no sentido do calendário, mas no sentido analítico. Por conseguinte, um período pode envolver uma determinada extensão de tempo de calendário em um modelo econômico específico, mas uma outra extensão totalmente diferente em outro modelo. Além disso, até no mesmo modelo não devemos interpretar que cada período sucessivo seja necessariamente igual ao tempo de calendário equivalente. No sentido analítico, um período é uma mera quantidade de tempo que transcorre antes de a variável y sofrer uma variação.

17.1 Tempo discreto, diferenças e equações de diferenças

A mudança de tempo contínuo para tempo discreto não produz nenhum efeito sobre a natureza fundamental da análise dinâmica, embora a formulação do problema deva ser alterada. Basicamente, nosso problema dinâmico ainda é encontrar uma trajetória temporal a partir de algum determinado padrão de variação da variável y ao longo do tempo. Mas, agora, o padrão de variação deve ser representado pelo quociente de diferenças $\Delta y / \Delta t$, que é a contraparte de tempo discreto da derivada dy/dt . Contudo, lembre-se que agora t só pode assumir valores inteiros; assim, quando estivermos comparando os valores de y em dois períodos consecutivos, devemos ter $\Delta t = 1$. Por essa razão, o quociente de diferenças $\Delta y / \Delta t$ pode ser simplificado para a expressão Δy ; essa expressão é denominada a *primeira diferença de y* . De acordo com isso, o símbolo Δ , que significa

diferença, pode ser interpretado como uma diretriz para tomar a primeira diferença de (y) . Como tal, constitui a contraparte de tempo discreto do símbolo de operador d/dt .

A expressão Δy pode assumir vários valores, dependendo, é claro, de quais dois períodos consecutivos de tempo estejam envolvidos na operação de diferença (ou “diferenciação”). Para evitar ambigüidade, vamos adicionar um índice de tempo a y e definir a primeira diferença mais especificamente como se segue:

$$\Delta y_t \equiv y_{t+1} - y_t \quad (17.1)$$

onde y_t significa o valor de y no t -ésimo período, e y_{t+1} é seu valor no período imediatamente após o t -ésimo período. Com essa simbologia, podemos descrever o padrão de mudança de y por uma equação como

$$\Delta y_t = 2 \quad (17.2)$$

ou

$$\Delta y_t = -0,1y_t \quad (17.3)$$

Equações desse tipo são denominadas *equações de diferenças*. Note a extraordinária semelhança entre as duas últimas equações, por um lado, e as equações diferenciais $dy/dt = 2$ e $dy/dt = -0,1y$, pelo outro.

Ainda que as equações de diferenças derivem seu nome de expressões de diferenças tal como Δy , há formas alternativas equivalentes dessas equações que são completamente livres de expressões Δ e que são mais convenientes de usar. Em virtude de (17.1), podemos reescrever (17.2) como

$$y_{t+1} - y_t = 2 \quad (17.2')$$

ou

$$y_{t+1} = y_t + 2 \quad (17.2'')$$

Para (17.3), as formas alternativas equivalentes correspondentes são

$$y_{t+1} - 0,9y_t = 0 \quad (17.3')$$

ou

$$y_{t+1} = 0,9y_t \quad (17.3'')$$

As versões numeradas identificadas por duas “linhas” serão convenientes quando estivermos calculando um valor de y a partir de um valor conhecido de y do período precedente. Todavia, em discussões posteriores, empregaremos, em grande parte, as versões numeradas identificadas com uma “linha”, isto é, as que aparecem em (17.2') e (17.3').

É importante notar que a escolha de índices de tempo em uma equação de diferenças é, de certa forma, arbitrária. Por exemplo, sem que haja nenhuma mudança de significado, (17.2') pode ser reescrita como $y_t - y_{t-1} = 2$, onde $(t-1)$ refere-se ao período que precede imediatamente o t -ésimo período. Ou podemos expressá-la, de modo equivalente, como $y_{t+2} - y_{t+1} = 2$.

Além disso, podemos salientar que, embora tenhamos usado consistentemente símbolos y com índices, também é aceitável usar $y(t)$, $y(t+1)$ e $y(t-1)$ em seu lugar. Contudo, de modo a evitar a utilização da notação $y(t)$ para ambos os casos de tempo contínuo e tempo discreto, vamos adotar o mecanismo de índices na discussão da análise de período.

Assim como as equações diferenciais, as equações de diferenças podem ser lineares ou não-lineares, homogêneas ou não-homogêneas e de primeira ou segunda ordem (ou de ordens

mais altas). Analise, por exemplo, (17.2'). Ela pode ser classificada como: (1) linear, pois nenhum termo y (de qualquer período) está elevado à segunda potência (ou à potência mais alta) ou está multiplicado por um termo y de um outro período; (2) não-homogênea, visto que o lado direito (onde não há nenhum termo y) é diferente de zero; e (3) de primeira ordem, porque existe somente uma *diferença primeira* Δy_t , que envolve somente uma defasagem de tempo de um período. (Por outro lado, uma equação de diferenças de segunda ordem, a ser discutida no Capítulo 18, envolve uma defasagem de dois períodos e, assim, acarreta três termos y : y_{t+2} , y_{t+1} , bem como y_t .)

Na verdade, (17.2') também pode ser caracterizada como tendo coeficientes constantes e um termo constante (= 2). Uma vez que o caso de coeficiente constante é o único que vamos considerar, daqui em diante admitiremos implicitamente essa caracterização. Por todo o presente capítulo, também será mantida a característica do termo constante, e no Capítulo 18 discutiremos um método para tratar o caso do termo variável.

Observe que a equação (17.3') também é linear e da primeira ordem; mas, diferentemente de (17.2'), ela é homogênea.

17.2 Resolvendo uma equação de diferenças de primeira ordem

Ao resolver uma equação diferencial, nosso objetivo era achar uma trajetória temporal $y(t)$. Como sabemos, tal trajetória temporal é uma função de tempo que é totalmente livre de quaisquer expressões de derivadas (ou diferenciais) e que é perfeitamente consistente com a equação diferencial dada, bem como com suas condições iniciais. A trajetória temporal que procuramos com uma equação de diferenças é de natureza semelhante. Novamente, ela deve ser uma função de t — uma fórmula que defina os valores de y em cada um dos períodos de tempo — e que seja consistente com a equação de diferenças dada, bem como com suas condições iniciais. Além do mais, ela não deve conter nenhuma expressão de diferenças, tal como Δy_t (ou expressões como $y_{t+1} - y_t$).

Resolver equações diferenciais é, em última instância, uma questão de integração. Como resolvemos uma equação de diferenças?

Método iterativo

Antes de desenvolver um método geral de ataque, em primeiro lugar vamos explicar um método relativamente prosaico, o *método iterativo* — que, embora grosseiro, provará ser imensamente revelador da natureza essencial de uma solução.

Neste capítulo, estamos preocupados somente com o caso de primeira ordem; assim, a equação de diferenças descreve o padrão de variação de y entre *dois* períodos consecutivos apenas. Uma vez especificado tal padrão, assim como por (17.2''), e uma vez que tenhamos um valor inicial y_0 , não é problema achar y_1 pela equação. De modo semelhante, uma vez encontrado y_1 , y_2 poderá ser obtido imediatamente, e assim por diante, pela aplicação repetida (iteração) do padrão de variação especificado na equação de diferenças. Os resultados da iteração então nos permitirão inferir uma trajetória temporal.

EXEMPLO 1

Encontre a solução da equação de diferenças (17.2), supondo um valor inicial de $y_0 = 15$. Para executar o processo iterativo, é mais conveniente usar a forma alternativa da equação de diferenças (17.2''), a saber, $y_{t+1} = y_t + 2$, com $y_0 = 15$. Por essa equação, podemos deduzir, passo a passo, que

$$\begin{aligned} y_1 &= y_0 + 2 \\ y_2 &= y_1 + 2 = (y_0 + 2) + 2 = y_0 + 2(2) \\ y_3 &= y_2 + 2 = [y_0 + 2(2)] + 2 = y_0 + 3(2) \\ &\dots \end{aligned}$$

e, em geral, para qualquer período t ,

$$y_t = y_0 + t(2) = 15 + 2t \quad (17.4)$$

Esta última equação indica o valor de y de qualquer período de tempo (incluindo o período inicial $t = 0$); por conseguinte, ela constitui a solução de (17.2).

O processo de iteração é grosseiro – corresponde aproximadamente a resolver equações diferenciais simples por integração direta –, mas serve para indicar claramente a maneira pela qual uma trajetória temporal é gerada. Em geral, o valor de y_t dependerá, de um modo especificado, do valor de y no período imediatamente anterior (y_{t-1}); assim, um dado valor inicial y_0 levará sucessivamente a $y_1, y_2, \dots$, por meio do padrão de variação prescrito.

EXEMPLO 2

Resolva a equação de diferenças (17.3); desta vez, deixe que o valor inicial não especificado seja denotado simplesmente por y_0 . Novamente, é mais conveniente trabalhar com a versão alternativa em (17.3''), a saber, $y_{t+1} = 0,9y_t$. Por iteração, temos

$$\begin{aligned} y_1 &= 0,9y_0 \\ y_2 &= 0,9y_1 = 0,9(0,9y_0) = (0,9)^2 y_0 \\ y_3 &= 0,9y_2 = 0,9(0,9)^2 y_0 = (0,9)^3 y_0 \\ &\dots \end{aligned}$$

Essas expressões podem ser resumidas na solução

$$y_t = (0,9)^t y_0 \quad (17.5)$$

Para despertar maior interesse, podemos acrescentar algum conteúdo econômico a este exemplo. Pela simples análise do multiplicador, um único dispêndio de investimento no período 0 exigirá sucessivas rodadas de gastos que, por sua vez, originarão quantidades variadas de incremento de receita em períodos de tempo sucessivos. Usando y para denotar *incremento de receita*, temos y_0 igual ao montante de investimento no período 0; mas os incrementos de receita subseqüentes dependerão da propensão marginal ao consumo (MPC). Se $MPC = 0,9$ e se a receita de cada período for consumida somente no período seguinte, então 90% de y_0 serão consumidos no período 1, resultando em um incremento de receita de $y_1 = 0,9y_0$ no período 1. Por raciocínio semelhante, podemos encontrar $y_2 = 0,9y_1$ etc. Como podemos ver, esses resultados são exatamente os resultados do processo iterativo citado anteriormente. Em outras palavras, o processo multiplicador de geração de receita pode ser descrito por uma equação de diferenças tal como (17.3''), e uma solução como (17.5) nos informará qual deverá ser a grandeza do incremento de receita em qualquer período de tempo t .

EXEMPLO 3

Resolva a equação de diferenças homogênea

$$my_{t+1} - ny_t = 0$$

Após normalização e transposição, essa expressão pode ser escrita como

$$y_{t+1} = \left(\frac{n}{m}\right) y_t$$

que é a mesma que (17.3'') no Exemplo 2, exceto pela substituição de 0,9 por n/m . Assim, por analogia, a solução deve ser

$$y_t = \left(\frac{n}{m}\right)^t y_0$$

Observe o termo $\left(\frac{n}{m}\right)^t$. É por meio desse termo que vários valores de t levarão a seus valores correspondentes de y . Por conseguinte, ele corresponde à expressão e^{rt} nas soluções de equações diferenciais. Se o escrevermos de modo mais geral como b^t (b de base) e acrescentarmos a cons-

tante multiplicativa mais geral A (em vez de y_0), vemos que a solução da equação de diferenças geral homogênea do Exemplo 3 será da forma

$$y_t = Ab^t$$

Veremos que essa expressão Ab^t desempenhará o mesmo papel importante em equações de diferenças que a expressão Ae^{rt} desempenhou em equações diferenciais.[†] Todavia, ainda que ambas sejam expressões exponenciais, a primeira é na base b , ao passo que a última é na base e . É óbvio que, exatamente como o tipo da trajetória temporal contínua $y(t)$ guarda uma grande dependência do valor de r , a trajetória temporal discreta y_t depende principalmente do valor de b .

Método geral

Agora, você já deve estar bastante impressionado com as várias semelhanças entre as equações diferenciais e as equações de diferenças. Como já era de se esperar, o método geral de solução que será explicado dentro em pouco será paralelo ao método para equações diferenciais.

Suponha que estamos procurando a solução para a equação de diferenças de primeira ordem

$$y_{t+1} + ay_t = c \quad (17.6)$$

onde a e c são duas constantes. A solução geral consistirá na soma de dois componentes: uma *solução particular* y_p , que é *qualquer* solução da equação completa não-homogênea (17.6), e uma *função complementar* y_c , que é a solução geral da equação homogênea associada a (17.6):

$$y_{t+1} + ay_t = 0 \quad (17.7)$$

O componente y_p novamente representa o nível de equilíbrio intertemporal de y , e o componente y_c os desvios entre a trajetória temporal e esse equilíbrio. A soma de y_c e y_p constitui a *solução geral*, por causa da presença de uma constante arbitrária. Como antes, para definir a solução, é preciso uma condição inicial.

Em primeiro lugar, vamos tratar da função complementar. Nossa experiência com o Exemplo 3 sugere que podemos tentar uma solução da forma $y_t = Ab^t$ (com $Ab^t \neq 0$, senão, y_t resultará simplesmente em uma reta horizontal sobre o eixo t); nesse caso, também temos $y_{t+1} = Ab^{t+1}$. Se esses valores de y_t e y_{t+1} forem válidos, a equação homogênea (17.7) se tornará

$$Ab^{t+1} + aAb^t = 0$$

a qual, com o cancelamento do fator comum diferente de zero Ab^t , resulta em

$$b + a = 0 \quad \text{ou} \quad b = -a$$

Isso significa que, para que a solução experimental funcione, temos de estabelecer $b = -a$; então a função complementar deve ser escrita como

$$y_c (= Ab^t) = A(-a)^t$$

Agora, vamos procurar a solução particular, que tem a ver com a equação completa (17.6). Nesse particular, o Exemplo 3 não auxilia em nada, porque refere-se apenas a uma equação homogênea. Contudo, notamos que, para y_p , podemos escolher *qualquer* solução de (17.6); assim, se uma solução experimental da forma mais simples $y_t = k$ (uma constante) puder resolver a questão, não encontraremos nenhuma dificuldade real. Agora, se $y_t = k$, então y manterá o mesmo va-

[†] Você talvez se oponha a esse enunciando ressaltando que a solução (17.4) no Exemplo 1 não contém um termo na forma de Ab^t . Esse fato, contudo, surge somente porque no Exemplo 1 temos $b = n/m = 1/1 = 1$, de modo que o termo Ab^t se reduz a uma constante.

lor constante ao longo do tempo, e deveremos ter $y_{t+1} = k$ também. A substituição desses valores em (17.6) resulta em

$$k + ak = c \quad \text{e} \quad k = \frac{c}{1+a}$$

Visto que esse valor particular de k satisfaz a equação, a solução particular pode ser escrita como

$$y_p(=k) = \frac{c}{1+a} \quad (a \neq -1)$$

Como isso é uma constante, um equilíbrio constante é indicado nesse caso.

Contudo, se por acaso $a = -1$, como no Exemplo 1, a solução particular $c/(1+a)$ não é definida, e alguma outra solução da equação não-homogênea (17.6) deve ser procurada. Nesse caso, empregamos o conhecido truque de tentar uma solução da forma $y_t = kt$. Isso implica, é claro, que $y_{t+1} = k(t+1)$. Substituindo essas expressões em (17.6), temos

$$k(t+1) + akt = c \quad \text{e} \quad k = \frac{c}{t+1+at} = c \quad [\text{porque } a = -1]$$

assim,

$$y_p(=kt) = ct$$

Essa forma da solução particular é uma função não-constante de t ; por conseguinte, ela representa um equilíbrio móvel.

Somando y_c com y_p , agora podemos escrever a solução geral em uma das duas formas seguintes:

$$y_t = A(-a)^t + \frac{c}{1+a} \quad [\text{solução geral, caso de } a \neq -1] \quad (17.8)$$

$$y_t = A(-a)^t + ct = A + ct \quad [\text{solução geral, caso de } a = -1] \quad (17.9)$$

Nenhuma dessas duas formas é completamente determinada, em vista da constante arbitrária A . Para eliminar essa constante arbitrária, recorreremos à condição inicial de que $y_t = y_0$ quando $t = 0$. Fazendo $t = 0$ em (17.8), temos

$$y_0 = A + \frac{c}{1+a} \quad \text{e} \quad A = y_0 - \frac{c}{1+a}$$

Por conseqüência, a versão definida de (17.8) é

$$y_t = \left(y_0 - \frac{c}{1+a} \right) (-a)^t + \frac{c}{1+a} \quad [\text{solução definida, caso de } a \neq -1] \quad (17.8')$$

Por outro lado, fazendo $t = 0$ em (17.9), obtemos $y_0 = A$, portanto, a versão definida de (17.9) é

$$y_t = y_0 + ct \quad [\text{solução definida, caso de } a = -1] \quad (17.9')$$

Se este último resultado for aplicado ao Exemplo 1, a solução que surge é exatamente a mesma que a solução iterativa (17.4).

Você pode verificar a validade de cada uma dessas soluções pelas duas etapas a seguir. Primeiro, fazendo $t = 0$ em (17.8'), observe que a última equação se reduz à identidade $y_0 = y_0$, o que significa que a condição inicial é satisfeita. Em segundo lugar, substituindo a fórmula (17.8') para y_t e uma fórmula semelhante para y_{t+1} – obtida pela substituição de t por $(t+1)$ em (17.8') – em (17.6),

vemos que a última se reduz à identidade $c = c$, o que significa que a trajetória temporal é consistente com a equação de diferenças dada. A verificação da validade da solução (17.9') é análoga.

EXEMPLO 4 Resolva a equação de diferenças de primeira ordem

$$y_{t+1} - 5y_t = 1 \quad \left(y_0 = \frac{7}{4}\right)$$

Seguindo o procedimento usado na dedução de (17.8'), podemos encontrar y_c experimentando uma solução $y_t = Ab^t$ (que implica $y_{t+1} = Ab^{t+1}$). Substituindo esses valores na versão homogênea $y_{t+1} - 5y_t = 0$ e cancelando o fator comum Ab^t , obtemos $b = 5$. Assim,

$$y_c = A(5)^t$$

Para encontrar y_p , tente a solução $y_t = k$, que implica $y_{t+1} = k$. Substituindo essas expressões na equação de diferenças completa, obtemos $k = -\frac{1}{4}$. Portanto,

$$y_p = -\frac{1}{4}$$

Segue-se que a solução geral é

$$y_t = y_c + y_p = A(5)^t - \frac{1}{4}$$

Fazendo $t = 0$ aqui e utilizando a condição inicial $y_0 = \frac{7}{4}$, obtemos $A = 2$. Desse modo, por fim, a solução definida pode ser escrita como

$$y_t = 2(5)^t - \frac{1}{4}$$

Visto que a equação de diferenças deste exemplo é um caso especial de (17.6), com $a = -5$, $c = 1$ e $y_0 = \frac{7}{4}$, e, uma vez que (17.8') é a "fórmula" de solução para esse tipo de equação de diferenças, poderíamos ter achado nossa solução inserindo os valores específicos dos parâmetros em (17.8'), obtendo como resultado

$$y_t = \left(\frac{7}{4} - \frac{1}{1-5}\right)(5)^t + \frac{1}{1-5} = 2(5)^t - \frac{1}{4}$$

que está perfeitamente de acordo com a resposta anterior.

Note que o termo y_{t+1} em (17.6) tem um coeficiente unitário. Se o coeficiente desse termo da equação de diferenças dada for diferente da unidade, ele deve ser normalizado antes de usarmos a fórmula de solução (17.8').

EXERCÍCIO 17.2

1. Converta as seguintes equações de diferenças para a forma de (17.2''):

(a) $\Delta y_t = 7$

(b) $\Delta y_t = 0,3y_t$

(c) $\Delta y_t = 2y_t - 9$

2. Resolva as seguintes equações de diferenças por iteração:

(a) $y_{t+1} = y_t + 1$ ($y_0 = 10$)

(b) $y_{t+1} = \alpha y_t$ ($y_0 = \beta$)

(c) $y_{t+1} = \alpha y_t - \beta$ ($y_t = y_0$ onde $t = 0$)

3. Reescreva as equações no Problema 2 na forma de (17.6), e resolva aplicando a fórmula (17.8') ou (17.9'), a que for mais apropriada. Suas respostas estão de acordo com as obtidas pelo método iterativo?
4. Para cada uma das seguintes equações de diferenças, use o procedimento ilustrado na dedução de (17.8') e (17.9') para encontrar y_c , y_p e a solução definida:

(a) $y_{t+1} + 3y_t = 4$	$(y_0 = 4)$
(b) $2y_{t+1} - y_t = 6$	$(y_0 = 7)$
(c) $y_{t+1} = 0,2y_t + 4$	$(y_0 = 4)$

17.3 A estabilidade dinâmica de equilíbrio

No caso do tempo contínuo, a estabilidade dinâmica de equilíbrio depende do termo Ae^{rt} na função complementar. Na análise de período, o papel correspondente é desempenhado pelo termo Ab^t na função complementar. Uma vez que sua interpretação é, de certa maneira, mais complicada do que Ae^{rt} , vamos tentar esclarecer antes de prosseguir.

A significado de b

O fato de o equilíbrio ser dinamicamente estável é uma questão de a função complementar tender ou não a zero à medida que $t \rightarrow \infty$. Basicamente, devemos analisar a trajetória do termo Ab^t quando t aumenta indefinidamente. É óbvio que o valor de b (a base desse termo exponencial) é de crucial importância nessa questão. Vamos considerar, em primeiro lugar, apenas seu significado, desprezando o coeficiente A (supondo que $A = 1$).

Para finalidades analíticas, podemos dividir a faixa de valores possíveis de b , $(-\infty, +\infty)$, em sete regiões distintas, como mostram as duas primeiras colunas da Tabela 17.1, arranjadas em ordem decrescente de grandeza de b . Essas regiões também são demarcadas na Figura 17.1 sobre uma escala vertical de b , sendo os pontos $+1$, 0 e -1 os pontos de demarcação. Na verdade, esses últimos três pontos constituem, por si, as regiões II, IV e VI. As regiões III e V, por outro lado, correspondem ao conjunto dos valores absolutos menores que a unidade que são, respectivamente, positivos e negativos. As duas regiões restantes, I e VII, são as regiões onde o valor absoluto de b é maior que a unidade.

Em cada região, a expressão exponencial b^t gera um tipo diferente de trajetória temporal. Esses tipos são exemplificados na Tabela 17.1 e ilustrados na Figura 17.1. Na região I (onde $b > 1$), b^t deve crescer com t a um passo crescente. Portanto, a configuração geral da trajetória temporal assumirá a forma do gráfico na parte superior da Figura 17.1. Note que esse gráfico é mostrado como uma função escalonada, e não como uma curva suave; isso porque estamos tratando com análise de período. Na região II ($b = 1$), b^t permanecerá na unidade para todos os valores de t . Por isso, seu gráfico será uma reta horizontal. Em seguida, na região III, b^t representa um número positivo menor que a unidade elevado a potências inteiras. À medida que a potência aumenta, b^t deve decrescer, embora permanecendo sempre positiva. O caso seguinte, de $b = 0$ na região IV, é bastante similar ao caso de $b = 1$; mas, aqui, temos $b^t = 0$ em vez de $b^t = 1$, portanto seu gráfico coincidirá com o eixo horizontal. Contudo, esse caso é de interesse periférico apenas, já que anteriormente adotamos a premissa de que $Ab^t \neq 0$.

Quando passamos para as regiões negativas, ocorre um novo fenômeno interessante: o valor de b^t se alterna entre valores positivos e negativos de período a período! Esse fato é claramente evidenciado nas três últimas linhas da Tabela 17.1 e nos três últimos gráficos da Figura 17.1. Na região V, onde b é um número negativo com módulo menor que 1, a trajetória temporal alternante tende a se aproximar cada vez mais do eixo horizontal (conforme a região positiva correspondente, III). Por outro lado, quando $b = -1$ (região VI), resulta uma alternância perpétua entre $+1$ e -1 . E, finalmente, quando $b < -1$ (região VII), a trajetória temporal alternante se afastará cada vez mais do eixo horizontal.

O surpreendente é que, conquanto não haja nenhuma possibilidade de o fenômeno de uma trajetória temporal flutuante se originar de um único termo Ae^{rt} (o caso da raiz complexa da equação diferencial de segunda ordem requer um par de raízes complexas), a flutuação pode ser gerada por um único termo b^t (ou Ab^t). Note, entretanto, que o caráter da flutuação é um pouco

				Valor de b^t em diferentes períodos de tempo				
Região	Valor de b		Valor de b^t	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4...$
I	$b > 1$	$(b > 1)$	por exemplo, $(2)^t$	1	2	4	8	16
II	$b = 1$	$(b = 1)$	$(1)^t$	1	1	1	1	1
III	$0 < b < 1$	$(b < 1)$	por exemplo, $(\frac{1}{2})^t$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
IV	$b = 0$	$(b = 0)$	$(0)^t$	0	0	0	0	0
V	$-1 < b < 0$	$(b < 1)$	por exemplo, $(-\frac{1}{2})^t$	1	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{8}$	$\frac{1}{16}$
VI	$b = -1$	$(b = 1)$	$(-1)^t$	1	-1	1	-1	1
VII	$b < -1$	$(b > 1)$	por exemplo, $(-2)^t$	1	-2	4	-8	16

TABELA 17.1
Uma classificação dos valores de b

diferente; ao contrário do padrão da função circular, a flutuação retratada na Figura 17.1 não é suave. Por essa razão, empregaremos a palavra *oscilação* para denotar o novo tipo de flutuação não-suave, ainda que muitos autores usem os termos flutuação e oscilação indiferentemente.

A essência da discussão precedente pode ser transmitida no seguinte enunciado geral: a trajetória temporal de b^t ($b \neq 0$) será

$$\left. \begin{array}{l} \text{Não-oscilatória} \\ \text{Oscilatória} \end{array} \right\} \text{ se } \left\{ \begin{array}{l} b > 0 \\ b < 0 \end{array} \right.$$

$$\left. \begin{array}{l} \text{Divergente} \\ \text{Convergente} \end{array} \right\} \text{ se } \left\{ \begin{array}{l} |b| > 1 \\ |b| < 1 \end{array} \right.$$

É importante notar que, ao passo que a convergência da expressão e^{rt} depende do *sinal* de r , a convergência da expressão b^t depende, por sua vez, do *valor absoluto* de b .

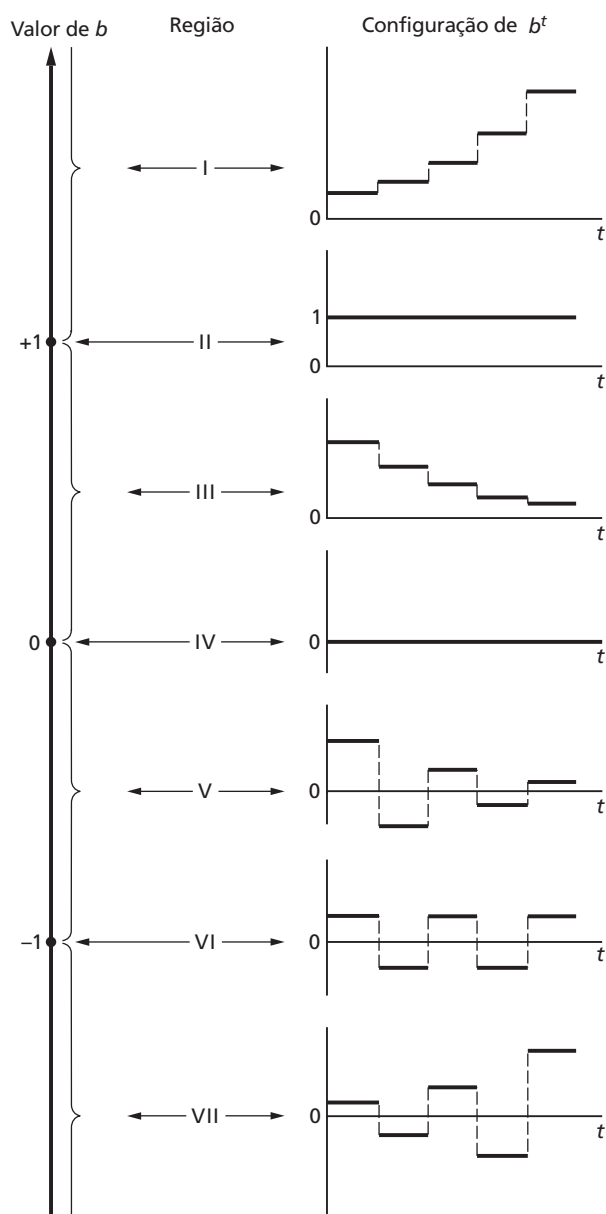
O papel de A

Até aqui, omitimos deliberadamente a constante multiplicativa A . Mas seus efeitos – e há dois deles – são relativamente fáceis de levar em conta. Em primeiro lugar, a *grandeza* de A pode servir para “ampliar” (se, digamos, $A = 3$) ou “reduzir” (se, digamos, $A = \frac{1}{5}$) os valores de b^t . Isto é, ela pode produzir um *efeito de escala* sem mudar a configuração básica da trajetória temporal. O *sinal* de A , por outro lado, afeta materialmente a forma da trajetória porque, se b^t for multiplicado por $A = -1$, então cada trajetória temporal mostrada na Figura 17.1 será substituída por sua própria imagem especular com referência ao eixo horizontal. Assim, um A negativo pode produzir um *efeito de espelho*, bem como um efeito de escala.

Convergência ao equilíbrio

A discussão anterior apresenta a interpretação do termo Ab^t na função complementar, a qual, como recordamos, representa os desvios em relação a algum nível de equilíbrio intertemporal. Se um termo (digamos) $y_p = 5$ for acrescentado ao termo Ab^t , a trajetória temporal deve ser deslocada na vertical por um valor constante de 5. Isso não afetará, de modo algum, a convergência ou a divergência da trajetória temporal, mas alterará o nível que serve como referência para a medi-

FIGURA 17.1



da da convergência ou da divergência. A Figura 17.1 retrata a convergência (ou falta dela) da expressão Ab^t em relação a zero. Quando y_p é incluído, passa a ser uma questão da convergência da trajetória temporal $y_t = y_c + y_p$ em relação ao nível de equilíbrio y_p .

Com relação a isso, vamos acrescentar algumas palavras de explicação para o caso especial de $b = 1$ (região II). Uma trajetória temporal tal como

$$y_t = A(1)^t + y_p = A + y_p$$

dá a impressão de convergir, porque o termo multiplicativo $(1)^t = 1$ não produz nenhum efeito explosivo. Observe, contudo, que y_t agora assumirá o valor $(A + y_p)$, e não o valor de equilíbrio y_p ; na verdade, ele nunca pode alcançar y_p (a menos que $A = 0$). Como uma ilustração desse tipo de situação, podemos citar a trajetória temporal em (17.9), na qual está envolvido um equilíbrio móvel $y_p = ct$. Essa trajetória temporal deve ser considerada divergente, não porque t aparece na solução particular, mas porque, com um A diferente de zero, haverá um desvio constante em relação ao equilíbrio móvel. Assim, ao estipular a condição para a convergência da trajetória temporal y_t ao equilíbrio y_p , devemos excluir o caso de $b = 1$.

Em suma, a solução

$$y_t = Ab^t + y_p$$

é uma trajetória convergente se, e somente se, $|b| < 1$.

EXEMPLO 1 Que tipo de trajetória temporal é representada por $y_t = 2\left(-\frac{4}{5}\right)^t + 9$? Visto que $b = -\frac{4}{5} < 0$, a trajetória temporal é oscilatória. Mas, uma vez que $|b| = \frac{4}{5} < 1$, a oscilação é amortecida e a trajetória temporal converge para o nível de equilíbrio de 9.

Você deve tomar cuidado para não confundir $2\left(-\frac{4}{5}\right)^t$ com $-2\left(\frac{4}{5}\right)^t$; eles representam configurações de trajetória temporal inteiramente diferentes.

EXEMPLO 2 Como se caracteriza a trajetória temporal $y_t = 3(2)^t + 4$? Uma vez que $b = 2 > 0$, não ocorrerá nenhuma oscilação. Mas, visto que $|b| = 2 > 1$, a trajetória temporal divergirá em relação ao nível de equilíbrio de 4.

EXERCÍCIO 17.3

1. Discuta a natureza das seguintes trajetórias temporais:

(a) $y_t = 3^t + 1$

(c) $y_t = 5\left(-\frac{1}{10}\right)^t + 3$

(b) $y_t = 2\left(\frac{1}{3}\right)^t$

(d) $y_t = -3\left(\frac{1}{4}\right)^t + 2$

2. Qual é a natureza da trajetória temporal obtida de cada equação de diferenças no Exercício 17.2–4?

3. Ache as soluções das seguintes e determine se as trajetórias temporais são oscilatórias e convergentes:

(a) $y_{t+1} - \frac{1}{3}y_t = 6 \quad (y_0 = 1)$

(c) $y_{t+1} + \frac{1}{4}y_t = 5 \quad (y_0 = 2)$

(b) $y_{t+1} + 2y_t = 9 \quad (y_0 = 4)$

(d) $y_{t+1} - y_t = 3 \quad (y_0 = 5)$

17.4 O modelo da teia de aranha

Para ilustrar a utilização de equações de diferenças de primeira ordem em análise econômica, citaremos duas variantes do modelo de mercado para uma única mercadoria. A primeira variante, conhecida como o modelo da *teia de aranha*, é diferente de nossos modelos de mercado anteriores porque trata Q_t não como uma função do preço corrente, mas do preço no período precedente.

O modelo

Considere uma situação na qual o produtor deve tomar a decisão de produção um período antes da venda real – tal como acontece na produção agrícola, onde o plantio deve anteceder a colheita e a venda da produção por um período de tempo apreciável. Vamos supor que a decisão de produção no período t é baseada no preço vigente à época, P_t . Contudo, uma vez que essa produção não estará disponível para a venda até um período $(t+1)$, P_t não determinará Q_{st} , mas $Q_{s,t+1}$. Assim, temos uma função oferta “defasada”.[†]

[†] Aqui, estamos adotando a premissa implícita de que toda a produção de um período será colocada no mercado, sem que nenhuma parte dela seja armazenada. Essa premissa é adequada quando a mercadoria em questão é perecível ou quando nunca se mantém estoque. Um modelo com estoque será considerado na Seção 17.5.

$$Q_{s,t+1} = S(P_t)$$

ou, equivalentemente, retroagindo os índices do tempo em um período,

$$Q_{st} = S(P_{t-1})$$

Quando essa função oferta interagir com uma função demanda da forma

$$Q_{dt} = D(P_t)$$

teremos como resultado interessantes padrões de preços dinâmicos.

Tomando as versões lineares dessas funções oferta (defasada) e demanda (não-defasada) e supondo que em cada período de tempo o preço de mercado é sempre estabelecido em um nível que compensa o mercado, temos um modelo de mercado com as três equações seguintes:

$$\begin{aligned} Q_{dt} &= Q_{st} \\ Q_{dt} &= \alpha - \beta P_t \quad (\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_{t-1} \quad (\gamma, \delta > 0) \end{aligned} \quad (17.10)$$

Substituindo as duas últimas equações na primeira, contudo, o modelo pode ser reduzido a uma única equação diferenças de primeira ordem, como se segue:

$$\beta P_t + \delta P_{t-1} = \alpha + \gamma$$

Para resolver essa equação, é desejável, em primeiro lugar, normalizá-la e deslocar os índices de tempo de para um período à frente [alterar t para $(t+1)$ etc.]. O resultado,

$$P_{t+1} + \frac{\delta}{\beta} P_t = \frac{\alpha + \gamma}{\beta} \quad (17.11)$$

então, será uma réplica de (17.6), com as substituições

$$y = P \quad a = \frac{\delta}{\beta} \quad e \quad c = \frac{\alpha + \gamma}{\beta}$$

Considerando que δ e β são ambos positivos, segue-se que $a \neq -1$. Por consequência, podemos aplicar a fórmula (17.8') para obter a trajetória temporal

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta} \right) \left(-\frac{\delta}{\beta} \right)^t + \frac{\alpha + \gamma}{\beta + \delta} \quad (17.12)$$

onde P_0 representa o preço inicial.

As teias de aranha

Três pontos podem ser observados no que se refere a essa trajetória temporal. Em primeiro lugar, a expressão $(\alpha + \gamma)/(\beta + \delta)$, que constitui a solução particular da equação de diferenças, pode ser considerada como o preço do equilíbrio intertemporal do modelo:[†]

$$\bar{P} = \frac{\alpha + \gamma}{\beta + \delta}$$

[†] No que concerne ao equilíbrio no sentido da compensação de mercado, o preço alcançado em cada período é um preço de equilíbrio, porque supusemos $Q_{dt} = Q_{st}$ para todo t .

Como isso é uma constante, o equilíbrio também é constante. Substituindo $\bar{P}$ em nossa solução, podemos expressar a trajetória temporal P_t alternativamente na forma

$$P_t = (P_0 - \bar{P}) \left(-\frac{\delta}{\beta} \right)^t + \bar{P} \quad (17.12')$$

Isso nos leva ao segundo ponto, a saber, o significado da expressão $(P_0 - \bar{P})$. Uma vez que essa expressão corresponde à constante A no termo Ab^t , seu sinal estará relacionado à questão de a trajetória temporal começar acima ou abaixo do equilíbrio (efeito espelho), ao passo que sua grandeza decidirá o quanto acima ou abaixo (efeito de escala) ela estará. Por último, há a expressão $(-\delta/\beta)$, que corresponde ao componente b de Ab^t . Como nosso modelo especifica que $\beta, \delta > 0$, podemos deduzir uma trajetória temporal oscilatória. É esse fato que dá origem ao fenômeno da teia de aranha, como veremos dentro em pouco. É claro que podem surgir *três* variedades possíveis de padrões de oscilação no modelo. Segundo a Tabela 17.1 ou Figura 17.1, a oscilação será

$$\left. \begin{array}{l} \text{Explosiva} \\ \text{Uniforme} \\ \text{Atenuada} \end{array} \right\} \text{ se } \delta \gtrless \beta$$

onde o termo *oscilação uniforme* refere-se ao tipo de trajetória na região VI.

Para visualizar as teias de aranha, vamos retratar o modelo (17.10) na Figura 17.2. A segunda equação de (17.10) é representada graficamente por uma curva linear de demanda com inclinação descendente numericamente igual a β . De modo semelhante, podemos desenhar uma curva linear de oferta com uma inclinação igual a δ a partir da terceira equação, se fizermos o eixo Q representar, nessa instância, uma quantidade ofertada *defasada*. O caso de $\delta > \beta$ (S mais inclinada do que D) e o caso de $\delta < \beta$ (S menos inclinada do que D) são ilustrados nas Figuras 17.2a e b, respectivamente. Em qualquer caso, entretanto, a interseção de D e S resultará no preço de equilíbrio intertemporal $\bar{P}$.

Quando $\delta > \beta$, como na Figura 17.2a, a interação entre demanda e oferta produzirá uma oscilação explosiva como se segue. Dado um preço inicial P_0 (que aqui admitimos estar acima de $\bar{P}$), podemos seguir a ponta da seta e ler na curva S que a quantidade ofertada no período seguinte (período 1) será Q_1 . Para compensar o mercado, a quantidade demandada no período 1 também deve ser Q_1 , o que é possível se, e somente se, o preço for estabelecido no nível de P_1 (ver seta dirigida para baixo). Agora, por meio da curva S , o preço P_1 levará a Q_2 como sendo a quantidade fornecida no período 2 e, para compensar o mercado nesse último período, o preço deve ser estabelecido no nível de P_2 , de acordo com a curva de demanda. Repetindo esse raciocínio, podemos traçar os preços e quantidades em períodos subseqüentes simplesmente seguindo as pontas das setas no diagrama e tecendo, desse modo, uma “teia de aranha” ao redor das curvas de

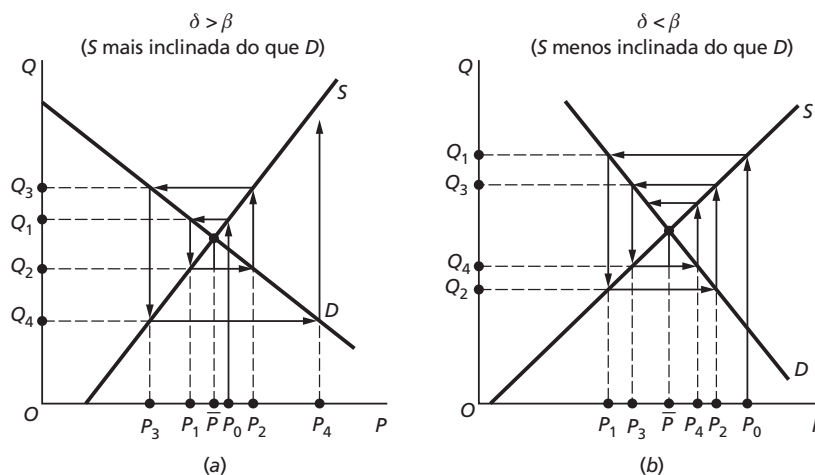


FIGURA 17.2

demanda e oferta. Comparando os níveis de preço, $P_0, P_1, P_2, \dots$, observamos, nesse caso, não somente um padrão oscilatório de variação, mas também uma tendência de ampliação do desvio do preço em relação a $\bar{P}$ à medida que o tempo passa. Como a teia de aranha é tecida de dentro para fora, a trajetória temporal é divergente e a oscilação é explosiva.

Por comparação, no caso da Figura 17.2b, onde $\delta < \beta$, o processo de tecelagem cria uma teia de aranha centrípeta. Se seguirmos as pontas das setas a partir de P_0 , nos aproximaremos cada vez mais da interseção entre as curvas de demanda e de oferta, onde $\bar{P}$ está situado. Embora ainda oscilatória, essa trajetória temporal é convergente.

Na Figura 17.2, não mostramos a terceira possibilidade, a saber, $\delta = \beta$. Todavia, o procedimento de análise gráfica envolvido é perfeitamente análogo aos dos outros dois casos. Por tanto, nós o deixamos para você, como exercício.

A discussão precedente tratou somente da trajetória temporal de P (isto é, P_t); contudo, após encontrar P_t , basta uma curta etapa para obter a trajetória temporal de Q . A segunda equação de (17.10) relaciona Q_{dt} com P_t , portanto, se (17.12) ou (17.12') for substituída na equação de demanda, a trajetória temporal de Q_{dt} pode ser obtida imediatamente. Além disso, uma vez que Q_{dt} deve ser igual a Q_{st} em cada período de tempo (compensação de mercado), podemos nos referir à trajetória temporal simplesmente como Q em vez de Q_{dt} . Com base na Figura 17.2, a idéia por trás dessa substituição pode ser vista com facilidade. Cada ponto sobre a curva D relaciona um P_t com uma Q pertencente ao mesmo período de tempo; por conseguinte, a função demanda pode servir para mapear a trajetória temporal do preço em relação à trajetória temporal da quantidade.

Você deve observar que a técnica gráfica da Figura 17.2 é aplicável mesmo quando as curvas D e S forem não-lineares.

EXERCÍCIO 17.4

- Com base em (17.10), encontre a trajetória temporal de Q e analise a condição para sua convergência.
- Desenhe um diagrama similar aos da Figura 17.2 para mostrar que, para o caso de $\delta = \beta$, o preço oscilará uniformemente sem amortecimento nem explosão.
- Dadas as seguintes demandas e ofertas para o modelo da teia de aranha, encontre o preço de equilíbrio intertemporal e determine se esse equilíbrio é estável:

(a) $Q_{dt} = 18 - 3P_t$	$Q_{st} = -3 + 4P_{t-1}$
(b) $Q_{dt} = 22 - 3P_t$	$Q_{st} = -2 + P_{t-1}$
(c) $Q_{dt} = 19 - 6P_t$	$Q_{st} = 6P_{t-1} - 5$
- No modelo (17.10), considere que a condição $Q_{dt} = Q_{st}$ e a função demanda fiquem como estão, mas mude a função oferta para

$$Q_{st} = -\gamma + \delta P_t^*$$
 onde P_t^* denota o *preço esperado* para o período t . Além do mais, suponha que a expectativa de preço dos vendedores é do tipo "adaptativo":[†]

$$P_t^* = P_{t-1}^* + \eta(P_{t-1} - P_{t-1}^*) \quad (0 < \eta \leq 1)$$
 onde η (a letra grega eta) é um coeficiente de ajuste de expectativa.
 - Dê uma interpretação econômica à equação precedente. Em que aspectos ela é semelhante e diferente da equação de expectativas adaptativas (16.34)?
 - O que acontece se η assumir seu valor máximo? Podemos considerar o modelo da teia de aranha como um caso especial do presente modelo?
 - Mostre que o novo modelo pode ser representado pela equação de diferenças de primeira ordem

[†] Ver Nerlove, Marc. "Adaptive Expectations and Cobweb Phenomena". *Quarterly Journal of Economics*, p. 227–240, maio 1958.

$$P_{t+1} - \left(1 - \eta - \frac{\eta\delta}{\beta}\right) P_t = \frac{\eta(\alpha + \gamma)}{\beta}$$

(Sugestão: Resolva a função oferta para P_t^* , e depois use a informação de que $Q_{st} = Q_{dt} = \alpha - \beta P_t$.)

- (d) Encontre a trajetória temporal do preço. Essa trajetória é necessariamente oscilatória? Ela pode ser oscilatória? Em que circunstâncias?
 - (e) Mostre que a trajetória temporal P_t se oscilatória, convergirá somente se $1 - 2/\eta < -\delta/\beta$. Em comparação com a solução da teia de aranha de (17.12) ou (17.12'), o novo modelo tem uma faixa mais ampla ou mais estreita para os valores indutores de estabilidade de $-\delta/\beta$?
5. O modelo da teia de aranha, assim como os modelos dinâmicos de mercado encontrados anteriormente, é essencialmente baseado no modelo estático de mercado apresentado na Seção 3.2. Qual premissa econômica é o agente dinamizador no presente caso? Explique.

17.5 Um modelo de mercado com estoque

No modelo precedente, admite-se que o preço é estabelecido de modo a compensar a produção corrente em cada um dos períodos de tempo. Essa premissa implica que ou a mercadoria é perecível e não pode ser estocada ou, embora ela possa ser estocada, nunca se mantém estoque. Agora, vamos construir um modelo no qual os vendedores mantêm um estoque da mercadoria.

O modelo

Vamos supor o seguinte:

1. A quantidade demandada, Q_{dt} , e a quantidade atualmente produzida, Q_{st} , são ambas funções lineares não defasadas do preço P_t .
2. O ajuste de preço não é efetuado por compensação de mercado em cada um dos períodos, mas por meio de um processo de estabelecimento de preços pelos vendedores: no início de cada período, eles estabelecem um preço para aquele período após levar em consideração a situação do estoque. Se, como resultado do preço do período precedente, houve acúmulo de estoque, o preço do período corrente será determinado em um nível mais baixo que o anterior, de modo a “movimentar” a mercadoria; mas se, ao contrário, for constatada uma baixa de estoque, o preço corrente será mais alto que o anterior.
3. O ajuste de preço feito de período a período é inversamente proporcional à variação de estoque observada.

Dadas essas premissas, podemos escrever as seguintes equações:

$$\begin{aligned} Q_{dt} &= \alpha - \beta P_t & (\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_t & (\gamma, \delta > 0) \\ P_{t+1} &= P_t - \sigma(Q_{st} - Q_{dt}) & (\sigma > 0) \end{aligned} \quad (17.13)$$

onde σ denota o coeficiente de *ajuste de preço induzido pelo estoque*. Note que, na verdade, (17.13) nada mais é do que a contraparte de tempo discreto do modelo de mercado da Seção 15.2, embora agora estejamos expressando o processo de ajuste de preço em termos de *estoque* ($Q_{st} - Q_{dt}$), e não de *excesso de demanda* ($Q_{dt} - Q_{st}$). Não obstante, os resultados analíticos revelarão ser muito diferentes; uma razão é que, com tempo discreto, podemos encontrar o fenômeno das oscilações. Vamos deduzir e analisar a trajetória temporal P_t .

A trajetória temporal

Substituindo as duas primeiras equações na terceira, o modelo pode ser condensado em uma única equação de diferenças:

$$P_{t+1} - [1 - \sigma(\beta + \delta)]P_t = \sigma(\alpha + \gamma) \quad (17.14)$$

e sua solução é dada por (17.8'):

$$\begin{aligned} P_t &= \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta} \right) [1 - \sigma(\beta + \delta)]^t + \frac{\alpha + \gamma}{\beta + \delta} \\ &= (P_0 - \bar{P}) [1 - \sigma(\beta + \delta)]^t + \bar{P} \end{aligned} \quad (17.15)$$

Portanto, é óbvio que a estabilidade dinâmica do modelo dependerá da expressão $1 - \sigma(\beta + \delta)$; por conveniência, vamos nos referir a essa expressão como b .

Com referência à Tabela 17.1, vemos que, ao analisar a expressão exponencial b^t , podem ser definidas sete regiões de valores de b . Contudo, uma vez que as especificações de nosso modelo ($\sigma, \beta, \delta > 0$) na verdade excluíram as duas primeiras regiões, restam apenas cinco casos possíveis, como relacionados na Tabela 17.2. Para cada uma dessas regiões, a especificação b da segunda coluna pode ser traduzida para uma especificação equivalente σ , como mostra a terceira coluna. Por exemplo, para a região III, a especificação de b é $0 < b < 1$; portanto, podemos escrever

$$0 < 1 - \sigma(\beta + \delta) < 1$$

$$-1 < -\sigma(\beta + \delta) < 0 \quad [\text{subtraindo 1 de todas as três partes}]$$

$$\text{e} \quad \frac{1}{\beta + \delta} > \sigma > 0 \quad [\text{dividindo tudo por } -(\beta + \delta)]$$

Esta última expressão nos dá a especificação equivalente desejada, σ , para a região III. A tradução para as outras regiões pode ser executada de modo análogo. Uma vez que o tipo de trajetória temporal pertinente a cada região já é conhecido pela Figura 17.1, a especificação σ nos habilita a dizer, pelos valores dados de σ, β , e δ , qual é a natureza geral da trajetória temporal P_t como delineado na última coluna da Tabela 17.2.

TABELA 17.2
Tipos de
trajetória
temporal

Região	Valor de $b \equiv 1 - \sigma(\beta + \delta)$	Valor de σ	Natureza da trajetória temporal P_t
III	$0 < b < 1$	$0 < \sigma < \frac{1}{\beta + \delta}$	Não oscilatória e convergente
IV	$b = 0$	$\sigma = \frac{1}{\beta + \delta}$	Permanece em equilíbrio [†]
V	$-1 < b < 0$	$\frac{1}{\beta + \delta} < \sigma < \frac{2}{\beta + \delta}$	Com oscilação amortecida
VI	$b = -1$	$\sigma = \frac{2}{\beta + \delta}$	Com oscilação uniforme
VII	$b < -1$	$\sigma = \frac{2}{\beta + \delta}$	Com oscilação explosiva

[†] O fato de o preço permanecer em equilíbrio nesse caso também pode ser verificado diretamente por (17.14). Com $\alpha = 1/(\beta + \delta)$, o coeficiente de P_t torna-se zero, e (17.14) se reduz a $P_{t+1} = \sigma(\alpha + \gamma) = (\alpha + \gamma)/(\beta + \delta) = \bar{P}$.

EXEMPLO 1

Se os vendedores de nosso modelo sempre aumentarem (reduzirem) o preço de 10% em relação à quantidade de redução (ou aumento) do estoque, e se a inclinação da curva de demanda for -1 e a da curva de oferta for 15 (ambas as inclinações referentes ao eixo do preço), que tipo de trajetória temporal P_t encontraremos?

Aqui, temos $\sigma = 0,1$, $\beta = 1$ e $\delta = 15$. Visto que $1/(\beta + \delta) = \frac{1}{16}$ e $2/(\beta + \delta) = \frac{1}{8}$, o valor de $\sigma (= \frac{1}{10})$ está localizado entre os dois valores anteriores; assim, é um caso da região V. A trajetória temporal P_t será caracterizada por oscilação amortecida.

Resumo gráfico dos resultados

A substância da Tabela 17.2, que contém cinco casos possíveis de especificação σ , pode ser muito mais fácil de compreender se os resultados forem apresentados graficamente. Considerando que a especificação σ envolve essencialmente uma comparação entre grandezas relativas dos parâmetros σ e $(\beta + \delta)$, vamos fazer o gráfico σ em função de $(\beta + \delta)$, como na Figura 17.3. Note que basta que nos preocupemos com o quadrante positivo porque, por especificação do modelo, σ e $(\beta + \delta)$ são ambos positivos. Pela Tabela 17.2, fica claro que as regiões IV e VI são especificadas, respectivamente, pelas equações $\sigma = 1/(\beta + \delta)$ e $\sigma = 2/(\beta + \delta)$. Visto que a representação gráfica de cada uma dessas equações é uma hipérbole retangular, as duas regiões são representadas graficamente pelas duas curvas hiperbólicas da Figura 17.3. Além disso, tão logo tenhamos as duas hipérboles, as outras três regiões se encaixam imediatamente no lugar. A região III, por exemplo, é o conjunto de pontos que ficam abaixo da hipérbole mais baixa, onde temos σ menor que $1/(\beta + \delta)$. De modo semelhante, a região V é representada pelo conjunto de pontos que fica entre as duas hipérboles, ao passo que todos os pontos localizados acima da hipérbole superior pertencem à região VII.

EXEMPLO 2

Se $\sigma = \frac{1}{2}$, $\beta = 1$ e $\delta = \frac{3}{2}$, nosso modelo (17.13) resultará em uma trajetória temporal convergente P_t ? Os valores paramétricos dados correspondem ao ponto A na Figura 17.3. Visto que o ponto está na região V, a trajetória temporal é convergente, embora oscilatória.

Você notará que, nos dois modelos que acabamos de apresentar, em cada instância nossos resultados analíticos são enunciados como um conjunto de casos alternativos possíveis – três tipos de trajetórias oscilatórias para as teias de aranha, e cinco tipos de trajetória temporal no modelo de estoque. Essa profusão de resultados analíticos se origina, é claro, da formulação paramétrica dos modelos. O fato de nosso resultado não poder ser enunciado em uma única resposta inequívoca é, obviamente, um mérito, e não uma desvantagem.

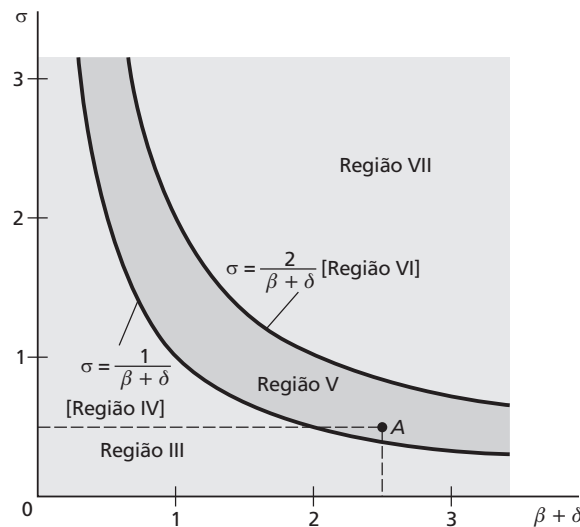


FIGURA 17.3

EXERCÍCIO 17.5

1. Na resolução de (17.14), por que deve ser usada a fórmula (17.8') em vez da (17.9')?
2. Tendo como base a Tabela 17.2, verifique a validade da tradução da especificação b para a especificação σ para as regiões IV a VII.
3. Se o modelo (17.13) tiver a seguinte forma numérica:

$$Q_{dt} = 21 - 2P_t$$

$$Q_{st} = -3 + 6P_t$$

$$P_{t+1} = P_t - 0,3(Q_{st} - Q_{dt})$$
 Encontre a trajetória temporal P_t e determine se ela é convergente.
4. Suponha que, no modelo (17.13), a oferta em cada período é uma quantidade fixa, digamos, $Q_{st} = k$, em vez de uma função do preço. Analise o comportamento do preço ao longo do tempo. Qual restrição deve ser imposta a k para tornar a solução economicamente significativa?

17.6 Equação de diferenças não-linear – a abordagem gráfico-qualitativa

Até aqui, utilizamos somente a equação de diferenças *linear* em nossos modelos; mas nem sempre os fatos da vida econômica concordam com a conveniência da linearidade. Felizmente, quando ocorre não-linearidade no caso de modelos de equações de diferenças de primeira ordem, existe um método fácil de análise que é aplicável sob condições razoavelmente gerais. Esse método, de natureza gráfica, é muito parecido com o da análise qualitativa de equações diferenciais de primeira ordem apresentado na Seção 15.6.

Diagrama de fase

Equações de diferenças não-lineares, nas quais aparecem somente as variáveis y_{t+1} e y_t , tais como

$$y_{t+1} + y_t^3 = 5 \quad \text{ou} \quad y_{t+1} + \sin y_t - \ln y_t = 3$$

podem ser representadas, em geral, pela equação

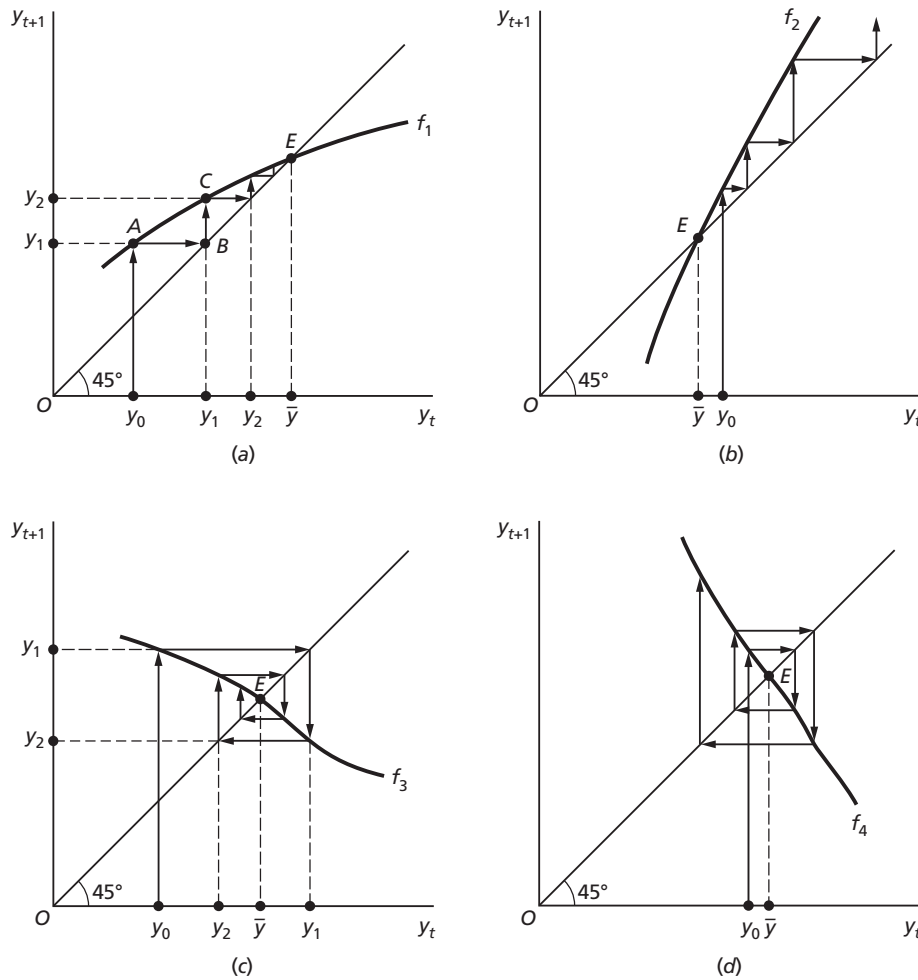
$$y_{t+1} = f(y_t) \quad (17.16)$$

onde f pode ser uma função de qualquer grau de complexidade, contanto que seja uma função apenas de y_t , sem t como um outro argumento. Quando fazemos o gráfico das duas variáveis y_{t+1} e y_t uma em relação à outra em um plano de coordenadas cartesianas, o diagrama resultante constitui um *diagrama de fase*, e a curva correspondente a f é uma *linha de fase*. Com isso, é possível analisar a trajetória temporal da variável pelo processo de iteração.

Os termos diagrama de fase e linha de fase são utilizados aqui de modo análogo ao do caso da equação diferencial; mas observe uma diferença na construção do diagrama. No caso da equação diferencial, desenhamos dy/dt em função de y como na Figura 15.3, de modo que, para que o presente caso seja perfeitamente análogo, devemos ter Δy_t no eixo vertical e y_t no eixo horizontal. Isso não é impossível de fazer, mas, em vez disso, é muito mais conveniente colocar y_{t+1} no eixo vertical, como fizemos na Figura 17.4, onde é usada a mesma escala para ambos os eixos. Note a presença de uma reta a 45° em cada diagrama da Figura 17.4; essa reta será muito útil na execução de nossa análise gráfica.

Vamos ilustrar o procedimento envolvido por meio da Figura 17.4a, onde desenhamos uma linha de fase (denominada f_1) que representa uma equação de diferenças específica $y_{t+1} = f_1(y_t)$. Se tivermos um valor inicial y_0 (desenhado no eixo horizontal), podemos traçar, por iteração, todos os valores subsequentes de y da seguinte maneira. Primeiro, visto que a linha de fase f_1 leva o valor inicial y_0 para y_1 de acordo com a equação

FIGURA 17.4



$$y_1 = f_1(y_0)$$

podemos ir diretamente de y_0 até a linha de fase, chegar ao ponto A e ler sua altura no eixo vertical como o valor de y_1 . Em seguida, procuramos mapear y_1 para y_2 de acordo com a equação

$$y_2 = f_1(y_1)$$

Para esse propósito, devemos primeiro desenhar y_1 no eixo horizontal – de modo semelhante a y_0 durante o primeiro mapeamento. Essa transposição necessária de y_1 do eixo vertical para o eixo horizontal é realizada com muita facilidade com a utilização da reta a 45° que, por ter uma inclinação de $+1$, é o lugar geométrico de pontos cujas abscissas e ordenadas são idênticas, tais como $(2, 2)$ e $(5, 5)$. Assim, para transpor y_1 do eixo vertical, podemos simplesmente atravessar a reta a 45° , chegar ao ponto B , e então seguir diretamente para baixo até o eixo horizontal para localizar o ponto y_1 . Repetindo esse processo, podemos ir de y_1 para y_2 via o ponto C na linha de fase e então usar a reta a 45° para transpor y_2 etc.

Agora que a natureza da iteração ficou clara, podemos observar que a iteração desejada pode ser alcançada simplesmente seguindo as setas de y_0 até A (na linha de fase), até B (na reta a 45°), até C (na linha de fase) etc. – sempre alternando entre as duas retas – sem jamais ser necessário recorrer novamente aos eixos.

Tipos de trajetória temporal

As iterações gráficas que acabamos de esboçar são, é claro, aplicáveis aos outros três diagramas na Figura 17.4. Na verdade, esses quatro diagramas servem para ilustrar quatro variedades básicas de linhas de fase, cada qual implicando um tipo diferente de trajetória temporal. As duas primeiras linhas de fase, f_1 e f_2 , são caracterizadas por inclinações positivas, sendo uma das inclinações menor que a unidade e a outra maior que a unidade:

$$0 < f_1'(y_t) < 1 \quad \text{e} \quad f_2'(y_t) > 1$$

As duas restantes, por outro lado, têm inclinações negativas; especificamente, temos

$$-1 < f_3'(y_t) < 0 \quad \text{e} \quad f_4'(y_t) < -1$$

Em cada diagrama da Figura 17.4, o valor do equilíbrio intertemporal de y (a saber, $\bar{y}$) está localizado na interseção da linha de fase com a reta a 45° , que denominamos E . Isso acontece porque o ponto E na linha de fase, por ser simultaneamente um ponto sobre a linha a 45° , levará um y_t a um y_{t+1} de valor idêntico; e quando $y_{t+1} = y_t$ por definição, y deve estar em equilíbrio intertemporal. Nossa tarefa principal é determinar se, dado um valor inicial de $y_0 \neq \bar{y}$, o padrão de variação implicado pela linha de fase nos levará consistentemente em direção a $\bar{y}$ (convergente) ou nos afastará dele (divergente).

Para a linha de fase f_1 , o processo iterativo leva de y_0 a $\bar{y}$ em uma trajetória sem oscilação. Você pode verificar que, se y_0 for colocado à direita de $\bar{y}$, haverá também um movimento em direção a $\bar{y}$, embora para a esquerda. Essas trajetórias temporais são convergentes ao equilíbrio e suas configurações gerais seriam do mesmo tipo das mostradas na região III da Figura 17.1.

Contudo, dada a linha de fase f_2 , cuja inclinação é maior que a unidade, surge uma trajetória temporal divergente. A partir de um valor inicial y_0 maior que $\bar{y}$, as setas se afastarão continuamente do equilíbrio em direção a valores de y cada vez mais altos. Como você pode verificar, um valor inicial mais baixo que $\bar{y}$ dá origem a um movimento continuamente divergente semelhante, embora na direção oposta.

Quando a inclinação da linha de fase for negativa, como em f_3 e f_4 , o movimento passa a ser oscilatório e agora aparece o fenômeno da *ultrapassagem* da marca de equilíbrio. No diagrama c , y_0 leva a y_1 , que é maior que $\bar{y}$, somente para ser seguido por y_2 , que fica abaixo de $\bar{y}$ etc. Nesses casos, a convergência da trajetória temporal dependerá de o valor absoluto da inclinação da linha de fase ser menor que 1. Esse é o caso da linha de fase f_3 , na qual a extensão da ultrapassagem tende a diminuir em períodos sucessivos. Por outro lado, para a linha de fase f_4 , cuja inclinação é maior que 1 em valor absoluto, prevalece a tendência oposta, resultando em uma trajetória temporal divergente.

As trajetórias temporais oscilatórias geradas pelas linhas de fase f_3 e f_4 lembram as teias de aranha na Figura 17.2. Na Figura 17.4c ou d , entretanto, a teia de aranha é tecida em torno de uma linha de fase (que contém uma defasagem) e da reta a 45° , em vez de ao redor de uma curva de demanda e de uma curva de oferta (defasada). Aqui, uma reta a 45° é usada como um auxílio mecânico para fazer a transposição de y , ao passo que, na Figura 17.2, a curva D (que desempenha um papel semelhante ao da reta a 45° grau na Figura 17.4) é uma parte integral do próprio modelo. Especificamente, uma vez determinada Q_{st} na curva de oferta, deixamos que as setas atinjam a curva D com o propósito de achar um preço que “compense o mercado”, como era a regra do jogo no modelo da teia de aranha. Por consequência, há uma diferença básica na denominação dos eixos: na Figura 17.2, há duas variáveis inteiramente diferentes, P e Q , mas, na Figura 17.4, os eixos representam os valores da mesma variável y em dois períodos consecutivos. Entretanto, note que, se analisarmos o gráfico da equação de diferenças (17.11), que resume o modelo da teia de aranha, em vez das funções demanda e oferta separadas em (17.10), o diagrama resultante será, então, uma linha de fase tal como mostra a Figura 17.4. Em outras palavras, realmente existem dois modos alternativos de analisar graficamente o modelo da teia de aranha, que darão o resultado idêntico.

A regra básica que surge dessa consideração da linha de fase é que o *sinal algébrico* de sua inclinação determina se haverá *oscilação* , e o *valor absoluto* de sua inclinação rege a questão da *convergência* . Se a linha de fase acaso contiver segmentos com inclinações positiva e negativa, e se o valor absoluto de sua inclinação for maior que 1 em alguns pontos e menor que 1 em outros, a trajetória temporal naturalmente se tornará mais complicada. Todavia, ainda em tais casos, a análise gráfico-iterativa pode ser empregada com igual facilidade. É claro que devemos dispor de um valor inicial antes de podermos iniciar a iteração. Na verdade, nesses casos mais complicados, um valor inicial diferente pode levar a uma trajetória temporal de uma espécie totalmente diferente (ver Exercícios 17.6–2 e 17.6–3).

Um mercado com um teto de preço

Agora citaremos um exemplo econômico de uma equação de diferenças não-linear. Na Figura 17.4, todas as quatro linhas de fase não-lineares são do tipo suave; no presente exemplo, mostraremos uma linha de fase não-suave.

Como ponto de partida, vamos tomar a equação de diferenças *linear* (17.11) do modelo da teia de aranha e reescrevê-la como

$$P_{t+1} = \frac{\alpha + \gamma}{\beta} - \frac{\delta}{\beta} P_t \quad \left(\frac{\delta}{\beta} > 0 \right) \quad (17.17)$$

Ela está no formato de $P_{t+1} = f(P_t)$, com $f'(P_t) = -\delta/\beta < 0$. Fazemos o gráfico dessa linha de fase linear na Figura 17.5 tendo como premissa que a inclinação é maior que a unidade em valor absoluto, o que implica uma oscilação *explosiva* .

Agora vamos impor um teto legal de preço $\hat{P}$ (lê-se “ *P* circunflexo” ou, menos formalmente, “ *P* chapéu”). Isso pode ser mostrado na Figura 17.5 como uma reta horizontal porque, agora, independentemente do nível de P_t , P_{t+1} não pode passar do nível de $\hat{P}$. Essa restrição invalida a parte da linha de fase que está acima de $\hat{P}$ ou, de uma perspectiva diferente, dobra para baixo a parte superior da linha de fase até o nível de $\hat{P}$, resultando, assim, em uma linha de fase quebrada.[†] Em vista dessa quebra, a nova linha de fase (traço mais escuro) não somente é não-linear, mas também é não-suave. Assim como uma função degrau, essa linha quebrada exigirá mais de uma equação para expressá-la algebricamente:

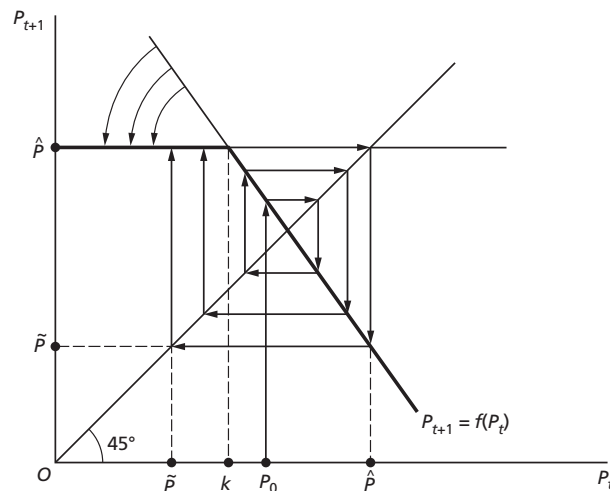


FIGURA 17.5

[†] Em termos estritos, deveríamos também “dobrar” a parte da linha de fase que está à direita do ponto $\hat{P}$ no eixo horizontal. Mas não faz mal algum deixar as coisas como estão, pois a transposição de P_{t+1} para o eixo horizontal já levará automaticamente o limite de $\hat{P}$ para o eixo P_t .

$$P_{t+1} = \begin{cases} \hat{P} & (\text{se } P_t \leq k) \\ \frac{\alpha + \gamma}{\beta} - \frac{\delta}{\beta} P_t & (\text{se } P_t > k) \end{cases} \quad (17.17')$$

onde k denota o valor de P_t no ponto de quebra.

Supondo um preço inicial P_0 , vamos traçar a trajetória temporal do preço iterativamente. Durante o primeiro estágio de iteração, quando estiver em efeito o segmento da linha de fase cuja inclinação é descendente, a tendência à oscilação explosiva manifesta-se claramente. Após alguns períodos, entretanto, as setas começam a atingir o teto de preço e, dali em diante, a trajetória temporal desenvolverá um movimento cíclico perpétuo entre $\hat{P}$ e um *piso efetivo de preço* $\tilde{P}$ (lê-se “ P til” ou, menos formalmente, “ P cobrinha”). Assim, em virtude do teto de preço, a tendência explosiva intrínseca do modelo é efetivamente contida e a oscilação, que era cada vez mais ampla, agora é amortecida até uma oscilação uniforme, produzindo o que é denominado um ciclo limite.

O aspecto significativo desse resultado é que, enquanto, no caso de uma linha de fase linear uma trajetória uniformemente oscilatória pode ser produzida se, e somente se, a inclinação da linha de fase for -1 , agora, após a introdução da não-linearidade, pode surgir o mesmo resultado mesmo quando a inclinação da linha de fase for diferente de -1 . A implicação econômica desse fato é de considerável importância. Se observarmos uma oscilação mais ou menos uniforme na trajetória temporal real de uma variável e tentarmos explicá-la por meio de um modelo *linear*, seremos forçados a recorrer à especificação de um modelo bastante especial – e implausível – no qual a inclinação da linha de fase é exatamente -1 . Mas, se a não-linearidade for introduzida, seja na variedade suave ou na variedade não-suave, então podemos usar uma profusão de premissas mais razoáveis, cada uma das quais pode ser responsável pelo aspecto observado de oscilação uniforme.

EXERCÍCIO 17.6

- Em modelos de equações de diferenças, a variável t pode assumir apenas valores inteiros. Isso implica que, nos diagramas de fase na Figura 17.4, as variáveis y_t e y_{t+1} devem ser consideradas variáveis discretas?
- Use, como linha de fase, a metade esquerda de uma curva em U invertida e deixe que ela intercepte a reta a 45° em dois pontos L (esquerda) e R (direita).
 - Esse é um caso de múltiplos equilíbrios?
 - Se o valor inicial y_0 estiver à esquerda de L , que tipo de trajetória temporal será obtido?
 - E se o valor inicial estiver entre L e R ?
 - E se o valor inicial estiver à direita de R ?
 - O que você pode concluir sobre a estabilidade dinâmica de equilíbrio em L e R , respectivamente?
- Use, como linha de fase, uma curva em U invertida. Deixe que seu segmento ascendente intercepte a reta a 45° no ponto L , e deixe que seu segmento descendente intercepte a reta a 45° no ponto R . Responda às mesmas cinco perguntas levantadas no Problema 2. (*Observação:* Sua resposta dependerá do modo particular como a linha de fase é desenhada; explore várias possibilidades.)
- Na Figura 17.5, abandone o teto legal de preço e imponha um preço mínimo P_m em seu lugar.
 - Qual será a mudança na linha de fase?
 - Ela será quebrada? Não-linear?
 - Haverá também o desenvolvimento de um movimento uniformemente oscilatório no preço?
- Com referência a (17.17') e Figura 17.5, mostre que a constante k pode ser expressa como

$$k = \frac{\alpha + \gamma}{\beta} - \frac{\beta}{\delta} \hat{P}$$

CAPÍTULO 18

Equações de diferenças de ordens mais altas

Os modelos econômicos no Capítulo 17 envolvem equações de diferenças que relacionam P_t e P_{t-1} entre si. Como o valor de P em um período pode determinar exclusivamente o valor de P no período seguinte, a trajetória temporal de P fica totalmente determinada tão logo seja especificado um valor inicial P_0 . Contudo, pode acontecer que o valor de uma variável econômica no período t (digamos, y_t) dependa não somente de y_{t-1} , mas também de y_{t-2} . Tal situação dará origem a uma equação de diferenças de segunda ordem.

Em termos estritos, uma *equação de diferenças de segunda ordem* é aquela que envolve uma expressão $\Delta^2 y_t$, denominada a *diferença segunda* de y_t , mas que não contém nenhuma diferença de ordem mais alta que 2. O símbolo Δ^2 , a contraparte de tempo discreto do símbolo d^2/dt^2 , é uma instrução para “tomar a diferença segunda” da seguinte maneira:

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_{t+1} - y_t) && [\text{por (17.1)}] \\ &= (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t) && [\text{novamente por (17.1)}]^\dagger \\ &= y_{t+2} - 2y_{t+1} + y_t\end{aligned}$$

Assim, a segunda diferença de y_t é transformável em uma soma de termos que envolve uma defasagem de tempo de dois períodos. Visto que expressões como $\Delta^2 y_t$ e Δy_t são bem incômodas para trabalhar, simplesmente redefiniremos uma equação de diferenças de segunda ordem como aquela que envolve uma defasagem de tempo de dois períodos na variável. De modo semelhante, uma equação de diferenças de terceira ordem é a que envolve uma defasagem de tempo de três períodos etc.

Em primeiro lugar, vamos nos concentrar no método de resolução de uma equação de diferenças de segunda ordem, deixando a generalização para equações de ordens mais altas na Seção 18.4. Para manter o escopo da discussão dentro de parâmetros manejáveis, neste capítulo trataremos apenas de equações de diferenças lineares com coeficientes constantes. Contudo, ambas as variedades, de termo constante e de termo variável, serão examinadas.

[†] Isto é, primeiro adiantamos os índices na expressão $(y_{t+1} - y_t)$ de um período, para obter uma nova expressão $(y_{t+2} - y_{t+1})$ e então subtraímos desta última a expressão original. Note que, uma vez que a diferença resultante pode ser escrita como $\Delta y_{t+1} - \Delta y_t$, podemos inferir a seguinte regra de operação:

$$\Delta(y_{t+1} - y_t) = \Delta y_{t+1} - \Delta y_t$$

Isso nos faz lembrar da regra aplicável à derivada de uma soma ou diferença.

18.1 Equações de diferenças lineares de segunda ordem com coeficientes constantes e termo constante

Uma variedade simples de equações de diferenças de segunda ordem toma a forma

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \quad (18.1)$$

Você reconhecerá que essa equação é linear, não-homogênea, com coeficientes constantes (a_1 , a_2) e termo constante c .

Solução particular

Como antes, pode-se esperar que a solução de (18.1) tenha dois componentes: uma solução particular y_p representando o nível de equilíbrio intertemporal de y , e uma função complementar y_c que especifica, para cada período de tempo, o desvio em relação ao equilíbrio. A solução particular, definida como qualquer solução da equação completa, às vezes pode ser obtida simplesmente tentando uma solução da forma $y_t = k$. Substituindo esse valor constante de y em (18.1), obtemos

$$k + a_1 k + a_2 k = c \quad \text{e} \quad k = \frac{c}{1 + a_1 + a_2}$$

Assim, contanto que $(1 + a_1 + a_2) \neq 0$, a solução particular é

$$y_p (= k) = \frac{c}{1 + a_1 + a_2} \quad (\text{caso de } a_1 + a_2 \neq -1) \quad (18.2)$$

EXEMPLO 1 Encontre a solução particular de $y_{t+2} - 3y_{t+1} + 4y_t = 6$. Aqui, temos $a_1 = -3$, $a_2 = 4$ e $c = 6$. Uma vez que $a_1 + a_2 \neq -1$, a solução particular pode ser obtida de (18.2) como se segue:

$$y_p = \frac{6}{1 - 3 + 4} = 3$$

Caso $a_1 + a_2 = -1$, então a solução experimental $y_t = k$ não funciona e podemos tentar $y_t = kt$ em seu lugar. Substituindo esta última em (18.1) e tendo em mente que agora temos $y_{t+1} = k(t+1)$ e $y_{t+2} = k(t+2)$, constatamos que

$$k(t+2) + a_1 k(t+1) + a_2 kt = c$$

$$\text{e} \quad k = \frac{c}{(1 + a_1 + a_2)t + a_1 + 2} = \frac{c}{a_1 + 2} \quad [\text{já que } a_1 + a_2 = -1]$$

Assim, podemos escrever a solução particular como

$$y_p (= kt) = \frac{c}{a_1 + 2} t \quad (\text{caso de } a_1 + a_2 = -1; a_1 \neq -2) \quad (18.2')$$

EXEMPLO 2 Encontre a solução particular de $y_{t+2} + y_{t+1} - 2y_t = 12$. Aqui, $a_1 = 1$, $a_2 = -2$ e $c = 12$. É óbvio que a fórmula (18.2) não é aplicável, mas (18.2') é. Assim,

$$y_p = \frac{12}{1 + 2} t = 4t$$

Esta solução particular representa um equilíbrio móvel.



Se $a_1 + a_2 = -1$, mas, ao mesmo tempo, $a_1 = -2$ (isto é, se $a_1 = -2$ e $a_2 = 1$), então podemos adotar uma solução experimental da forma $y_t = kt^2$, que implica $y_{t+1} = k(t+1)^2$ etc. Como você pode verificar, neste caso, a solução particular vem a ser

$$y_p = kt^2 = \frac{c}{2} t^2 \quad (\text{caso de } a_1 = -2; a_2 = 1) \quad (18.2'')$$

Contudo, visto que essa fórmula se aplica somente ao caso único da equação de diferenças $y_{t+2} - 2y_{t+1} + y_t = c$, sua utilidade é bastante limitada.

Função complementar

Para encontrar a função complementar, devemos nos concentrar na equação homogênea

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = 0 \quad (18.3)$$

Nossa experiência com equações de diferenças de primeira ordem nos ensinou que a expressão Ab^t desempenha um papel proeminente na solução geral de tal equação. Portanto, vamos tentar uma solução da forma $y_t = Ab^t$, a qual implica naturalmente que $y_{t+1} = Ab^{t+1}$, e assim por diante. Agora, nossa tarefa é determinar os valores de A e b .

Com a substituição da solução experimental em (18.3), a equação se torna

$$Ab^{t+2} + a_1 Ab^{t+1} + a_2 Ab^t = 0$$

ou, após cancelar o fator comum Ab^t (diferente de zero),

$$b^2 + a_1 b + a_2 = 0 \quad (18.3')$$

Essa equação quadrática – a *equação característica* de (18.3) ou de (18.1) –, que é comparável com (16.4''), possui as duas *raízes características*

$$b_1, b_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \quad (18.4)$$

cada uma das quais é aceitável na solução Ab^t . De fato, *ambas*, b_1 e b_2 , devem aparecer na solução geral da equação de diferenças homogênea (18.3) porque, exatamente como no caso de equações diferenciais, essa solução geral deve consistir em duas partes *linearmente independentes*, cada uma com sua própria constante multiplicativa arbitrária.

No que diz respeito às raízes características, podem ser encontradas três situações possíveis dependendo da expressão da raiz quadrada em (18.4). Você verá que isso guarda um paralelo muito próximo com a análise de equações diferenciais de segunda ordem na Seção 16.1.

Caso 1 (raízes reais distintas) Quando $a_1^2 > 4a_2$, a raiz quadrada em (18.4) é um número real e b_1 e b_2 são reais e distintas. Nesse caso, b_1^t e b_2^t são linearmente independentes e a função complementar pode ser escrita simplesmente como uma combinação linear dessas expressões, isto é,

$$y_c = A_1 b_1^t + A_2 b_2^t \quad (18.5)$$

Você deve comparar isso com (16.7).

EXEMPLO 3

Encontre a solução de $y_{t+2} + y_{t+1} - 2y_t = 12$. Essa equação tem os coeficientes $a_1 = 1$ e $a_2 = -2$; por (18.4), podemos constatar que as raízes características são $b_1, b_2 = 1, -2$. Assim, a função complementar é

$$y_c = A_1(1)^t + A_2(-2)^t = A_1 + A_2(-2)^t$$

Uma vez que no Exemplo 2 já constatamos que a solução particular da equação de diferenças dada é $y_p = 4t$, podemos escrever a solução geral como

$$y_t = y_c + y_p = A_1 + A_2(-2)^t + 4t$$

Ainda há duas constantes arbitrárias A_1 e A_2 a definir; para fazer isso, são necessárias *duas* condições iniciais. Suponha que temos $y_0 = 4$ e $y_1 = 5$. Então, uma vez que fazendo $t = 0$ e $t = 1$ sucessivamente na solução geral, encontramos

$$y_0 = A_1 + A_2 \quad (= 4 \text{ pela primeira condição inicial})$$

$$y_1 = A_1 - 2A_2 + 4 \quad (= 5 \text{ pela segunda condição inicial})$$

as constantes arbitrárias podem ser definidas como $A_1 = 3$ e $A_2 = 1$. Então, finalmente, a solução definida pode ser escrita como

$$y_t = 3 + (-2)^t + 4t$$

Caso 2 (raízes reais repetidas) Quando $a_1^2 = 4a_2$, a raiz quadrada em (18.4) se anula e as raízes características são repetidas:

$$b(= b_1 = b_2) = -\frac{a_1}{2}$$

Agora, se expressarmos a função complementar na forma de (18.5), os dois componentes passarão a ser um único termo:

$$A_1 b_1^t + A_2 b_2^t = (A_1 + A_2) b^t \equiv A_3 b^t$$

Isso não serve, porque agora temos uma constante a menos.

Para fornecer o componente que falta – o qual, é bom lembrar, deve ser linearmente independente do termo $A_3 b^t$ – mais uma vez o velho truque de multiplicar b^t pela variável t funcionará. Por conseguinte, o novo termo componente deve assumir a forma $A_4 t b^t$. É óbvio que ele é linearmente independente de $A_3 b^t$, pois nunca poderíamos obter a expressão $A_4 t b^t$ acrescentando um novo coeficiente a $A_3 b^t$. É fácil verificar que $A_4 t b^t$ realmente se qualifica como uma solução da equação homogênea (18.3), exatamente como $A_3 b^t$, substituindo $y_t = A_4 t b^t$ [e $y_{t+1} = A_4(t+1)b^{t+1}$ etc.] em (18.3)[†] e observando que a última expressão ficará reduzida a uma identidade $0 = 0$.

Por conseguinte, a função complementar para o caso de raízes repetidas é

$$y_c = A_3 b^t + A_4 t b^t \quad (18.6)$$

que você deve comparar com (16.9).

EXEMPLO 4 Encontre a função complementar de $y_{t+2} + 6y_{t+1} + 9y_t = 4$. Como os coeficientes são $a_1 = 6$ e $a_2 = 9$, constatamos que as raízes características são $b_1 = b_2 = -3$. Portanto, temos

$$y_c = A_3(-3)^t + A_4 t(-3)^t$$

Se prosseguirmos mais uma etapa, obtemos facilmente $y_p = \frac{1}{4}$, portanto a solução geral da equação de diferenças dada é

[†] Nessa substituição, devemos ter em mente que temos, no presente caso, $a_1^2 = 4a_2$ e $b = -a_1/2$.

$$y_t = A_3(-3)^t + A_4t(-3)^t + \frac{1}{4}$$

Dadas duas condições iniciais, novamente é possível atribuir valores definidos a A_3 e A_4 .

Caso 3 (raízes complexas) Segundo a possibilidade restante, $a_1^2 < 4a_2$, as raízes características são complexas conjugadas. Especificamente, elas serão da forma

$$b_1, b_2 = h \pm vi$$

onde

$$h = -\frac{a_1}{2} \quad \text{e} \quad v = \frac{\sqrt{4a_2 - a_1^2}}{2} \quad (18.7)$$

Assim, a função complementar em si torna-se

$$y_c = A_1b_1^t + A_2b_2^t = A_1(h + vi)^t + A_2(h - vi)^t$$

Do modo como se apresenta, y_c não é interpretada com facilidade. Felizmente, graças ao teorema de De Moivre, dado em (16.23'), essa função complementar pode ser facilmente transformada em termos trigonométricos, os quais já aprendemos a interpretar.

De acordo com o teorema citado, podemos escrever

$$(h \pm vi)^t = R^t(\cos \theta t \pm i \sin \theta t)$$

onde o valor de R (sempre tomado como positivo) é, por (16.10),

$$R = \sqrt{h^2 + v^2} = \sqrt{\frac{a_1^2 + 4a_2 - a_1^2}{4}} = \sqrt{a_2} \quad (18.8)$$

e θ é a medida do ângulo em radianos no intervalo $[0, 2\pi)$, que satisfaz as condições

$$\cos \theta = \frac{h}{R} = \frac{-a_1}{2\sqrt{a_2}} \quad \text{e} \quad \sin \theta = \frac{v}{R} = \sqrt{1 - \frac{a_1^2}{4a_2}} \quad (18.9)$$

Por conseguinte, a função complementar pode ser transformada da seguinte maneira:

$$\begin{aligned} y_c &= A_1R^t(\cos \theta t + i \sin \theta t) + A_2R^t(\cos \theta t - i \sin \theta t) \\ &= R^t[(A_1 + A_2) \cos \theta t + (A_1 - A_2)i \sin \theta t] \\ &= R^t(A_5 \cos \theta t + A_6 \sin \theta t) \end{aligned} \quad (18.10)$$

onde adotamos os símbolos abreviados

$$A_5 \equiv A_1 + A_2 \quad \text{e} \quad A_6 \equiv (A_1 - A_2)i$$

A função complementar (18.10) é diferente de sua contraparte, a equação diferencial (16.24'), em dois aspectos importantes. Primeiro, as expressões $\cos \theta t$ e $\sin \theta t$ substituíram as expressões usadas anteriormente, $\cos vt$ e $\sin vt$. Em segundo lugar, o fator multiplicativo R^t (uma exponencial de base R) substituiu a expressão exponencial natural e^{ht} . Em suma, passamos das coordenadas cartesianas (h e v) das raízes complexas para suas coordenadas polares (R e θ). Os valores de R e θ podem ser determinados por (18.8) e (18.9) tão logo h e v sejam conhecidos. Também é possível calcular R e θ diretamente dos valores dos parâmetros a_1 e a_2 via (18.8) e (18.9), contanto que primeiramente nos certifiquemos de que $a_1^2 < 4a_2$ e de que as raízes são, de fato, complexas.

EXEMPLO 5

Encontre a solução geral de $y_{t+2} + \frac{1}{4}y_t = 5$. Com coeficientes $a_1 = 0$ e $a_2 = \frac{1}{4}$, essa expressão constitui uma ilustração do caso de raiz complexa de $a_1^2 < 4a_2$. Por (18.7), as partes reais e imaginárias das raízes são $h = 0$ e $v = \frac{1}{2}$. Deduz-se, por (18.8), que

$$R = \sqrt{0 + \left(\frac{1}{2}\right)^2} = \frac{1}{2}$$

Uma vez que o valor de θ é o valor que pode satisfazer as duas equações

$$\cos \theta = \frac{h}{R} = 0 \quad \text{e} \quad \sin \theta = \frac{v}{R} = 1$$

podemos concluir, pela Tabela 16.1, que

$$\theta = \frac{\pi}{2}$$

Por consequência, a função complementar é

$$y_c = \left(\frac{1}{2}\right)^t \left(A_5 \cos \frac{\pi}{2}t + A_6 \sin \frac{\pi}{2}t \right)$$

Para encontrar y_p vamos tentar uma solução constante $y_t = k$ na equação completa, o que dá como resultado $k = 4$; assim, $y_p = 4$, e a solução geral pode ser escrita como

$$y_t = \left(\frac{1}{2}\right)^t \left(A_5 \cos \frac{\pi}{2}t + A_6 \sin \frac{\pi}{2}t \right) + 4 \quad (18.11)$$

EXEMPLO 6

Encontre a solução geral de $y_{t+2} - 4y_{t+1} + 16y_t = 0$. Em primeiro lugar, é fácil constatar que a solução particular é $y_p = 0$. Isso significa que a solução geral $y_t (= y_c + y_p)$ será idêntica a y_c . Para achar esta última, notamos que os coeficientes $a_1 = -4$ e $a_2 = 16$ de fato produzem raízes complexas. Assim, podemos substituir os valores de a_1 e a_2 diretamente em (18.8) e (18.9) para obter

$$R = \sqrt{16} = 4$$

$$\cos \theta = \frac{4}{2 \cdot 4} = \frac{1}{2} \quad \text{e} \quad \sin \theta = \sqrt{1 - \frac{16}{4 \cdot 16}} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

As duas últimas equações nos habilitam a concluir, pela Tabela 16.2, que

$$\theta = \frac{\pi}{3}$$

Se segue que a função complementar – que aqui também serve como a solução geral – é

$$y_c (= y_t) = 4^t \left(A_5 \cos \frac{\pi}{3}t + A_6 \sin \frac{\pi}{3}t \right) \quad (18.12)$$

A convergência da trajetória temporal

Assim como no caso de equações de diferenças de primeira ordem, a convergência da trajetória temporal y_t depende exclusivamente de y_c tender a zero quando $t \rightarrow \infty$. Por conseguinte, aquilo que aprendemos sobre as várias configurações da expressão b' na Figura 17.1 ainda é aplicável, embora no presente contexto tenhamos de considerar *duas* raízes características em vez de uma.

Considere, em primeiro lugar, o caso de raízes reais distintas: $b_1 \neq b_2$. Se $|b_1| > 1$ e $|b_2| > 1$, então, ambos os termos componentes na função complementar (18.5) – $A_1 b_1^t$ e $A_2 b_2^t$ – serão explosivos e, assim, y_c deve ser divergente. No caso oposto de $|b_1| < 1$ e $|b_2| < 1$, ambos os termos em y_c convergirão a zero à medida que t aumenta indefinidamente, o mesmo acontecendo com y_c . E se $|b_1| > 1$ mas $|b_2| < 1$? Nesse caso intermediário, é evidente que o termo $A_2 b_2^t$ tende a “definhar” enquanto o outro termo tende a se desviar cada vez mais de zero. Deduz-se, então, que o termo $A_1 b_1^t$ acabará dominando a cena e converterá a trajetória em divergente.



Vamos denominar *raiz dominante* a raiz que tenha o valor *absoluto* mais alto. Então é evidente que a raiz dominante b_1 é que realmente dá o tom da trajetória temporal, ao menos no que diz respeito à sua convergência ou divergência final. E é isso mesmo que acontece. Assim, podemos declarar que *uma trajetória temporal será convergente – sejam quais forem as condições iniciais – se, e somente se, a raiz dominante for menor que 1 em valor absoluto*. Você pode verificar que essa declaração é válida para os casos em que ambas as raízes são maiores ou menores que 1 em valor absoluto (discutido anteriormente) e uma das raízes tem valor absoluto exatamente igual a 1 (*não* discutido anteriormente). Entretanto, note que, embora a eventual convergência dependa apenas da raiz dominante, a raiz *não*-dominante também exercerá uma influência definida sobre a trajetória temporal, ao menos nos períodos iniciais. Por conseguinte, a exata configuração de y_t ainda depende de ambas as raízes.

Analisando o caso das raízes repetidas, constatamos que a função complementar consiste nos termos $A_3 b^t$ e $A_4 t b^t$, como mostra (18.6). O primeiro já nos é familiar, mas ainda precisamos dizer algumas palavras de explicação sobre o último, que envolve um t multiplicativo. Se $|b| > 1$, o termo b^t será explosivo e o t multiplicativo servirá simplesmente para intensificar essa característica à medida que t cresce. Se, por outro lado, $|b| < 1$, a parte b^t (que tende a zero à medida que t cresce) e a parte t se oporão uma à outra; isto é, o valor de t compensará, em vez de reforçar, o valor de b^t . Qual das forças se revelará a mais forte? A resposta é que a força de atenuação de b^t sempre vencerá a força de explosão de t . Por essa razão, o requisito básico para convergência no caso de raízes repetidas ainda é que a raiz seja menor que 1 em valor absoluto.

EXEMPLO 7 Analise a convergência das soluções nos Exemplos 3 e 4. Para o Exemplo 3, a solução é

$$y_t = 3 + (-2)^t + 4t$$

onde as raízes são 1 e -2 , respectivamente [$3(1)^t = 3$] e onde há um equilíbrio móvel $4t$. Como a raiz dominante é -2 , a trajetória temporal é divergente.

Para o Exemplo 4, onde a solução é

$$y_t = A_3(-3)^t + A_4 t(-3)^t + \frac{1}{4}$$

e onde $|b| = 3$, também temos divergência.

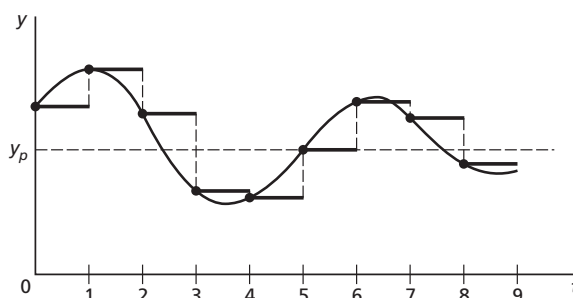
Agora, vamos considerar o caso da raiz complexa. Pela forma geral da função complementar em (18.10),

$$y_c = R^t (A_5 \cos \theta t + A_6 \sin \theta t)$$

fica claro que a expressão entre parênteses, assim como a expressão em (16.24'), produzirá um padrão flutuante de natureza periódica. Contudo, visto que a variável t só pode assumir valores inteiros 0, 1, 2, ... no presente contexto, vamos pegar e utilizar somente um subconjunto dos pontos no gráfico de uma função circular. O valor de y em cada um desses pontos sempre prevalecerá durante todo um período, até que seja alcançado o próximo ponto relevante. Como ilustrado na Figura 18.1, a trajetória resultante não é nem do tipo oscilatório usual (não se alterna entre valores acima e abaixo de y_p em períodos consecutivos), nem do tipo flutuante usual (não-suave); ao contrário, ela apresenta uma espécie de flutuação *degrau*. No que diz respeito à convergência, entretanto, o fator decisivo é realmente o termo R^t , o qual, assim como o termo e^{ht} em (16.24'), determinará se a flutuação degrau será intensificada ou amortecida à medida que t cresce. No presente caso, a flutuação pode ser gradualmente amortecida se, e somente se, $R < 1$. Uma vez que R é, por definição, o valor absoluto das raízes complexas conjugadas ($h \pm vi$), a condição para convergência é, novamente, que as raízes características sejam menores que a unidade em valor absoluto.

Resumindo: para todos os três casos de raízes características, a trajetória temporal convergirá para um equilíbrio intertemporal (constante ou móvel) – independentemente de quais sejam as condições iniciais – se, e somente se, o valor absoluto de cada uma das raízes for menor que 1.

FIGURA 18.1

**EXEMPLO 8**

As trajetórias temporais (18.11) e (18.12) são convergentes? Em (18.11), temos $R = \frac{1}{2}$; por conseguinte, a trajetória temporal convergirá para o equilíbrio constante ($= 4$). Em (18.12), por outro lado, temos $R = 4$, portanto a trajetória não convergirá para o equilíbrio ($= 0$).

EXERCÍCIO 18.1

- Escreva a equação característica para cada uma das seguintes e ache as raízes características:

(a) $y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 2$

(c) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 5$

(b) $y_{t+2} - 4y_{t+1} + 4y_t = 7$

(d) $y_{t+2} - 2y_{t+1} + 3y_t = 4$

- Para cada uma das equações de diferenças no Problema 1, determine, tendo como base suas raízes características, se a trajetória temporal envolve oscilação ou flutuação de grau, e se ela é explosiva.
- Encontre as soluções particulares das equações no Problema 1. Elas representam equilíbrios constantes ou móveis?
- Resolva as seguintes equações de diferenças:

(a) $y_{t+2} + 3y_{t+1} - \frac{7}{4}y_t = 9$

($y_0 = 6; y_1 = 3$)

(b) $y_{t+2} - 2y_{t+1} + 2y_t = 1$

($y_0 = 3; y_1 = 4$)

(c) $y_{t+2} - y_{t+1} + \frac{1}{4}y_t = 2$

($y_0 = 4; y_1 = 7$)

- Analise as trajetórias temporais obtidas no Problema 4.

18.2 Modelo de Samuelson para a interação multiplicador-aceleração

Como uma ilustração da utilização de equações de diferenças de segunda ordem na economia, vamos citar um trabalho clássico do professor Paul Samuelson, o primeiro economista a ganhar o Prêmio Nobel. Estamos nos referindo a seu clássico modelo de *interação*, que procura explorar o processo dinâmico de determinação de renda quando o princípio da aceleração está em ação juntamente com o multiplicador keynesiano.[†] Entre outras coisas, esse modelo serve para demonstrar que a mera interação entre o multiplicador e o acelerador é capaz de gerar flutuações cíclicas endogenamente.

[†] Samuelson, Paul A. "Interactions between the Multiplier Analysis and the Principle of Acceleration". *Review of Economic Statistics*, p. 75–78, maio 1939; reimpresso em *American Economic Association, Readings in Business Cycle Theory*. Homewood, Ill. Richard D. Irwin, Inc.: 1944, p. 261–269.

A estrutura

Suponha que a renda nacional Y_t é constituída de três correntes de dispêndio componentes: consumo C_t , investimento I_t e despesa do governo G_t . O consumo é visto como uma função não da renda corrente, mas da renda do período anterior, Y_{t-1} ; por simplicidade, supomos que C_t é estritamente proporcional a Y_{t-1} . O investimento, que é do tipo “induzido”, é uma função da tendência dominante de dispêndio do consumidor. É claro que é por meio desse investimento induzido que o princípio da aceleração entra no modelo. Especificamente, vamos supor que a razão entre I_t e o incremento de consumo $\Delta C_{t-1} \equiv C_t - C_{t-1}$ é fixa. Por outro lado, o terceiro componente, G_t , é considerado exógeno; na verdade, suporemos que ele é uma constante e o denotaremos simplesmente por G_0 .

Essas premissas podem ser traduzidas no seguinte conjunto de equações:

$$\begin{aligned} Y_t &= C_t + I_t + G_0 & (18.13) \\ C_t &= \gamma Y_{t-1} & (0 < \gamma < 1) \\ I_t &= \alpha (C_t - C_{t-1}) & (\alpha > 0) \end{aligned}$$

onde γ (a letra grega gama) representa a propensão marginal para consumir e α representa o acelerador (abreviatura de *coeficiente de aceleração*). Note que, se o investimento induzido for expurgado do modelo, ficaremos com uma equação de diferenças de primeira ordem que incorpora o processo do multiplicador dinâmico (cf. Exemplo 2 da Seção 17.2). Contudo, incluindo o investimento induzido, temos uma equação de diferenças de segunda ordem que retrata a interação entre o multiplicador e o acelerador.

Em virtude da segunda equação, podemos expressar I_t em termos de renda como se segue:

$$I_t = \alpha (\gamma Y_{t-1} - \gamma Y_{t-2}) = \alpha \gamma (Y_{t-1} - Y_{t-2})$$

Substituindo essa expressão e a equação C_t na primeira equação em (18.13) e rearranjando, o modelo pode ser condensado na única equação

$$Y_t - \gamma(1 + \alpha)Y_{t-1} + \alpha\gamma Y_{t-2} = G_0$$

ou, o que é equivalente (após adiantar os índices por dois períodos),

$$Y_{t+2} - \gamma(1 + \alpha)Y_{t+1} + \alpha\gamma Y_t = G_0 \quad (18.14)$$

Como essa é uma equação de diferenças linear de segunda ordem com coeficientes constantes e um termo constante, ela pode ser resolvida pelo método que acabamos de aprender.

A solução

Como solução particular, temos, por (18.2),

$$Y_p = \frac{G_0}{1 - \gamma(1 + \alpha) + \alpha\gamma} = \frac{G_0}{1 - \gamma}$$

Pode-se notar que a expressão $1/(1 - \gamma)$ é meramente o multiplicador que prevaleceria na ausência de investimento induzido. Assim, $G_0/(1 - \gamma)$ – o item exógeno referente à despesa vezes o multiplicador – deve nos dar a renda de equilíbrio Y^* no sentido de que esse nível de renda satisfaz a condição de equilíbrio “renda nacional = dispêndio total” [cf. (3.24)]. Entretanto, sendo a solução particular do modelo, ele também nos dá a renda de equilíbrio intertemporal $\bar{Y}$.

No que diz respeito à função complementar, há três casos possíveis. O Caso 1 ($a_1^2 > 4a_2$), no presente contexto, é caracterizado por

$$\gamma^2(1 + \alpha)^2 > 4\alpha\gamma \quad \text{ou} \quad \gamma(1 + \alpha)^2 > 4\alpha$$

ou

$$\gamma > \frac{4\alpha}{(1+\alpha)^2}$$

De modo semelhante, para caracterizar os Casos 2 e 3, basta mudar o sinal $>$ na última desigualdade para $=$ e $<$, respectivamente. Na Figura 18.2, desenhamos o gráfico da equação $\gamma = 4\alpha/(1+\alpha)^2$. De acordo com a discussão precedente, os pares (α, γ) que estão localizados exatamente *sobre* a curva pertencem ao Caso 2. Por outro lado, os pares (α, γ) localizados *acima* dessa curva (envolvendo valores mais altos de γ) têm a ver com o Caso 1, e os que estão localizados *abaixo* da curva têm a ver com o Caso 3.

Essa classificação tripartite, cuja representação gráfica é apresentada na Figura 18.2, é de interesse porque revela claramente as condições sob as quais podem emergir flutuações cíclicas endogenamente, advindas da interação entre o multiplicador e o acelerador. Mas isso nada revela sobre a convergência ou divergência da trajetória temporal de Y . Por conseguinte, ainda nos resta distinguir, em cada caso, entre os subcasos *amortecido* e *explosivo*. É claro que poderíamos optar pela saída fácil, que seria simplesmente ilustrar tais subcasos citando exemplos numéricos específicos. Porém, vamos tentar a tarefa mais recompensadora, se bem que mais trabalhosa, que é delinear as condições gerais sob as quais prevalecerão a convergência e a divergência.

Convergência versus divergência

A equação de diferenças (18.14) tem a equação característica

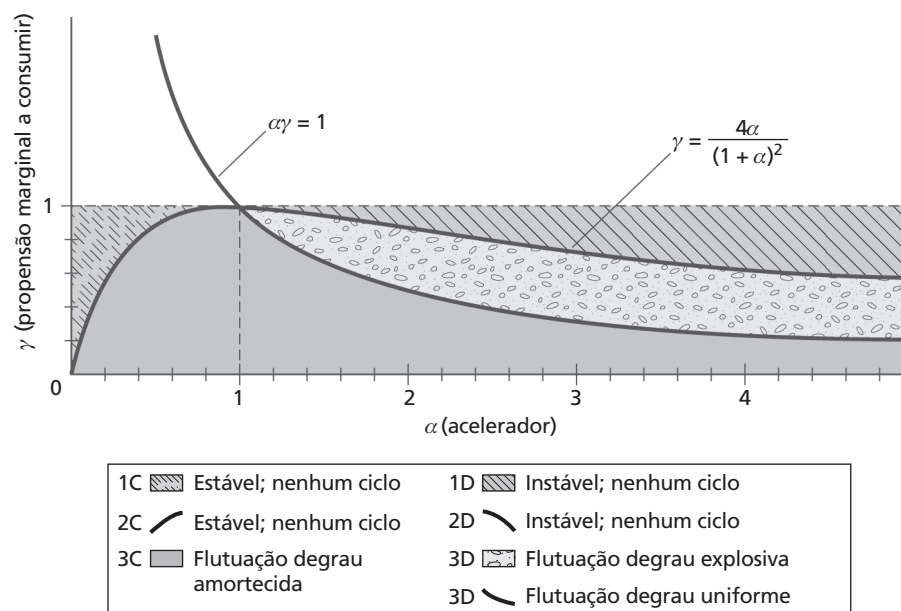
$$b^2 - \gamma(1+\alpha)b + \alpha\gamma = 0$$

que dá as duas raízes

$$b_1, b_2 = \frac{\gamma(1+\alpha) \pm \sqrt{\gamma^2(1+\alpha)^2 - 4\alpha\gamma}}{2}$$

Uma vez que a questão de convergência *versus* divergência depende dos valores de b_1 e b_2 e visto que b_1 e b_2 , por sua vez, dependem dos valores dos parâmetros α e γ , as condições para convergência e divergência devem poder ser expressas em termos dos valores de α e γ . Para fazer isso,

FIGURA 18.2



podemos recorrer ao fato de que – por (16.6) – as duas raízes características são sempre relacionadas uma com a outra pelas duas equações seguintes:

$$b_1 + b_2 = \gamma(1 + \alpha) \quad (18.15)$$

$$b_1 b_2 = \alpha\gamma \quad (18.15')$$

Com base nessas duas equações, podemos observar que

$$\begin{aligned} (1 - b_1)(1 - b_2) &= 1 - (b_1 + b_2) + b_1 b_2 \\ &= 1 - \gamma(1 + \alpha) + \alpha\gamma = 1 - \gamma \end{aligned} \quad (18.16)$$

Uma vez que o modelo especifica que $0 < \gamma < 1$, torna-se necessário impor às duas raízes a condição

$$0 < (1 - b_1)(1 - b_2) < 1 \quad (18.17)$$

Agora, vamos examinar a questão da convergência para o Caso 1, no qual as raízes são reais e distintas. Uma vez que, por hipótese, α e γ são ambos positivos, (18.15') nos diz que $b_1 b_2 > 0$, o que implica que b_1 e b_2 têm o mesmo sinal algébrico. Além do mais, visto que $\gamma(1 + \alpha) > 0$, (18.15) indica que b_1 e b_2 devem ser ambas positivas. Por conseguinte, a trajetória temporal Y_t não pode ter oscilações no Caso 1.

Ainda que agora os sinais de b_1 e b_2 sejam conhecidos, na verdade existem cinco combinações possíveis de valores de (b_1, b_2) para o Caso 1, cada uma delas com sua própria implicação no que diz respeito aos valores correspondentes para α e γ :

$$(i) \quad 0 < b_2 < b_1 < 1 \Rightarrow 0 < \gamma < 1; \alpha\gamma < 1$$

$$(ii) \quad 0 < b_2 < b_1 = 1 \Rightarrow \gamma = 1$$

$$(iii) \quad 0 < b_2 < 1 < b_1 \Rightarrow \gamma > 1$$

$$(iv) \quad 1 = b_2 < b_1 \Rightarrow \gamma = 1$$

$$(v) \quad 1 < b_2 < b_1 \Rightarrow 0 < \gamma < 1; \alpha\gamma > 1$$

A possibilidade i , na qual b_1 e b_2 são ambos números positivos menores que 1, satisfaz devidamente a condição (18.17) e, por conseguinte, está conforme a especificação do modelo $0 < \gamma < 1$. O produto das duas raízes também deve ser positivo e menor que 1 sob essa possibilidade e, por (18.15'), isso implica que $\alpha\gamma < 1$. Ao contrário, todas as três possibilidades seguintes violam a condição (18.17) e resultam em valores inadmissíveis de γ (veja Exercício 18.2-3). Por conseguinte, elas devem ser descartadas. Mas a Possibilidade v ainda pode ser aceitável. Quando b_1 e b_2 são ambas maiores que um, (18.17) ainda pode ser satisfeita se $(1 - b_1)(1 - b_2) < 1$. Mas dessa vez temos, por (18.15'), $\alpha\gamma > 1$ (e não < 1). O resultado final é que há somente dois subcasos admissíveis para o Caso 1. O primeiro – Possibilidade i – envolve raízes positivas menores que 1 b_1 e b_2 e, portanto, resulta em uma trajetória temporal convergente de Y . O outro subcaso – Possibilidade v – apresenta raízes maiores que um e, assim, produz uma trajetória temporal divergente. No que diz respeito aos valores de α e γ , todavia, a questão da convergência e da divergência depende somente de $\alpha\gamma < 1$ ou $\alpha\gamma > 1$. Essa informação está resumida na parte superior da Tabela 18.1, onde o subcaso convergente é denominado 1C, e o subcaso divergente, 1D.

A análise do Caso 2, com raízes repetidas, é semelhante. Agora, as raízes são $b = \gamma(1 + \alpha)/2$, com um sinal positivo porque α e γ são positivos. Assim, novamente não há nenhuma oscilação. Dessa vez podemos classificar o valor de b em apenas três possibilidades:

$$(vi) \quad 0 < b < 1 \quad \Rightarrow \quad \gamma < 1; \alpha\gamma < 1$$

$$(vii) \quad b = 1 \quad \Rightarrow \quad \gamma = 1$$

$$(viii) \quad b > 1 \quad \Rightarrow \quad \gamma < 1; \alpha\gamma > 1$$

Sob a Possibilidade *vi*, $b (= b_1 = b_2)$ é positivo e menor que 1; assim, as implicações referentes a α e γ são totalmente idênticas às da Possibilidade *i* para o Caso 1. Por analogia, a Possibilidade *viii*, com $b (= b_1 = b_2)$ maior que um, pode satisfazer (18.17) somente se $1 < b < 2$; nesse caso, ela leva aos mesmos resultados da Possibilidade *v*. Por outro lado, a Possibilidade *vii* viola (18.17) e deve ser descartada. Assim, novamente há apenas dois subcasos admissíveis. O primeiro – Possibilidade *vi* – resulta em uma trajetória temporal convergente, ao passo que o outro – Possibilidade *viii* – dá uma trajetória divergente. Em termos de α e γ , os subcasos convergente e divergente novamente estão associados com $\alpha\gamma < 1$ e $\alpha\gamma > 1$, respectivamente. Esses resultados estão relacionados na parte do meio da Tabela 18.1, onde os dois subcasos são denominados 2C (convergente) e 2D (divergente).

Por fim, no Caso 3, com raízes complexas, temos flutuação degrau e, por conseguinte, ciclos de negócios endógenos. Nesse caso, devemos observar o valor absoluto de $R = \sqrt{a_2}$ [veja (18.8)], que nos dará a orientação quanto à convergência e à divergência, onde a_2 é o coeficiente do termo y_t na equação de diferenças (18.1). No presente modelo, temos $R = \sqrt{\alpha\gamma}$, o que dá origem às três possibilidades seguintes:

$$(ix) \quad R < 1 \quad \Rightarrow \quad \alpha\gamma < 1$$

$$(x) \quad R = 1 \quad \Rightarrow \quad \alpha\gamma = 1$$

$$(xi) \quad R > 1 \quad \Rightarrow \quad \alpha\gamma > 1$$

Ainda que todas essas possibilidades sejam admissíveis (veja Exercício 18.2-4), somente a possibilidade $R < 1$ acarreta uma trajetória temporal convergente e se qualifica como Subcaso 3C na Tabela 18.1. Assim, as outras duas são denominadas coletivamente subcaso 3D.

Resumindo, podemos concluir, pela Tabela 18.1, que uma trajetória temporal convergente pode ocorrer se, e somente se, $\alpha\gamma < 1$.

Um resumo gráfico

A análise precedente resultou em uma classificação um tanto complexa de casos e subcasos. Uma representação visual do esquema de classificação seria um grande auxílio e é isso que nos dá a Figura 18.2.

TABELA 18.1
Casos e subcasos do Modelo de Samuelson

Caso	Subcaso	Valores de α e γ	Trajетória temporal Y_t
1. Raízes reais distintas $\gamma > \frac{4\alpha}{(1+\alpha)^2}$	1C: $0 < b_2 < b_1 < 1$ 1D: $1 < b_2 < b_1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	Não-oscilatória e não-flutuante
2. Raízes reais repetidas $\gamma = \frac{4\alpha}{(1+\alpha)^2}$	2C: $0 < b < 1$ 2D: $b > 1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	Não-oscilatória e não-flutuante
3. Raízes complexas $\gamma < \frac{4\alpha}{(1+\alpha)^2}$	3C: $R < 1$ 3D: $R \geq 1$	$\alpha\gamma < 1$ $\alpha\gamma \geq 1$	Com flutuação degrau

O conjunto de todos os pares (α, γ) admissíveis no modelo é mostrado na Figura 18.2 pela área retangular à qual foram aplicados diversos tipos de sombreamento. Uma vez que os valores de $\gamma = 0$ e $\gamma = 1$ são excluídos, assim como o valor de $\alpha = 0$, a área sombreada é uma espécie de retângulo sem lados. A equação $\gamma = 4\alpha/(1 + \alpha)^2$ já está representada graficamente na figura para demarcar os três casos principais da Tabela 18.1: os pontos sobre essa curva pertencem ao Caso 2; os pontos localizados ao norte da curva (que representam valores mais altos de γ) pertencem ao Caso 1; os que estão localizados ao sul (com valores menores de γ) são do Caso 3. Para distinguir entre os subcasos convergente e divergente, adicionamos agora o gráfico de $\alpha\gamma = 1$ (uma hipérbole retangular) para representar uma outra linha de demarcação. Os pontos que estão ao norte dessa hipérbole retangular satisfazem a desigualdade $\alpha\gamma > 1$, ao passo que os que estão localizados abaixo dela correspondem a $\alpha\gamma < 1$. Então é possível demarcar com facilidade os subcasos. Para o Caso 1, a região sombreada marcada com linhas interrompidas, por estar abaixo da hipérbole, corresponde ao Subcaso 1C, mas a região sombreada marcada por linhas contínuas está associada ao Subcaso 1D. Para o Caso 2, que está relacionado aos pontos situados sobre a curva $\gamma = 4\alpha/(1 + \alpha)^2$, o Subcaso 2C abrange a porção daquela curva cuja inclinação é ascendente e o Subcaso 2D, a porção cuja inclinação é descendente. Por fim, para o Caso 3, a hipérbole retangular serve para separar a região sombreada cinza (Subcaso 3C) da região sombreada marcada com pedrinhas (Subcaso 3D). Esta última, você deve anotar, inclui os pontos localizados *sobre* a própria hipérbole retangular, por causa da *desigualdade fraca* na especificação $\alpha\gamma \geq 1$.

Uma vez que a Figura 18.2 contém todas as conclusões qualitativas do modelo, dado qualquer par ordenado (α, γ) , sempre podemos encontrar o subcaso correto graficamente colocando o par ordenado no diagrama.

EXEMPLO 1

Se o acelerador for 0,8 e a propensão marginal a consumir for 0,7, que tipo de trajetória temporal de interação resultará? O par ordenado $(0,8; 0,7)$ está localizado na região sombreada cinza, Subcaso 3C; assim, a trajetória temporal é caracterizada por flutuação amortecida.

EXEMPLO 2

Que tipo de interação $\alpha = 2$ e $\gamma = 0,5$ implica? O par ordenado $(2; 0,5)$ está localizado exatamente sobre a hipérbole retangular, portanto, Subcaso 3D. A trajetória temporal de Y novamente apresentará flutuação degrau, mas não será nem explosiva nem amortecida. Por analogia com os casos de oscilação uniforme e flutuação uniforme, podemos denominar essa situação “flutuação degrau uniforme”. Contudo, em geral não se pode esperar que o aspecto de uniformidade neste último caso seja perfeito porque, semelhantemente ao que foi feito na Figura 18.1, só podemos aceitar os pontos sobre uma curva de seno ou de co-seno que correspondam a valores inteiros de t , mas esses valores de t podem atingir um conjunto inteiramente diferente de pontos sobre a curva em cada período de flutuação.

EXERCÍCIO 18.2

- Consultando a Figura 18.2, encontre os subcasos aos quais pertencem os conjuntos de valores α e γ e descreva a trajetória temporal da interação em termos qualitativos.
 - $\alpha = 3,5; \gamma = 0,8$
 - $\alpha = 2; \gamma = 0,7$
 - $\alpha = 0,2; \gamma = 0,9$
 - $\alpha = 1,5; \gamma = 0,6$
- Utilizando os valores de α e γ dados nas partes (a) e (c) do Problema 1, encontre os valores numéricos das raízes características em cada instância, e analise a natureza da trajetória temporal. Seus resultados estão de acordo com os obtidos anteriormente?
- Verifique se as Possibilidades ii, iii e iv para o Caso 1 implicam valores inadmissíveis de γ .
- Mostre que, no Caso 3, nunca podemos encontrar $\gamma \geq 1$.

18.3 Inflação e desemprego em tempo discreto

A interação entre inflação e desemprego, discutida anteriormente na estrutura do tempo contínuo, também pode ser enunciada para tempo discreto. Usando essencialmente as mesmas premissas econômicas, ilustraremos, nesta seção, como aquele modelo pode ser reformulado como um modelo de equação de diferenças.

O modelo

A formulação de tempo contínuo que aparece na Seção 16.5 consistia em três equações diferenciais:

$$p = \alpha - T - \beta U + g\pi \quad \begin{array}{l} \text{[relação de Philips com} \\ \text{expectativas aumentadas]} \end{array} \quad (16.33)$$

$$\frac{d\pi}{dt} = j(p - \pi) \quad \begin{array}{l} \text{[expectativas adaptativas]} \end{array} \quad (16.34)$$

$$\frac{dU}{dt} = -k(m - p) \quad \begin{array}{l} \text{[política monetária]} \end{array} \quad (16.35)$$

Estão presentes três variáveis endógenas: p (taxa de inflação real), π (taxa de inflação esperada) e U (taxa de desemprego). Seis parâmetros aparecem no modelo; entre eles está o parâmetro m – a taxa de crescimento da moeda nominal (ou a taxa de expansão monetária) –, que é diferente dos outros no sentido de que sua grandeza é estabelecida por uma decisão política.

Quando usamos o molde da análise de período, a relação de Phillips se torna, simplesmente,

$$p_t = \alpha - T - \beta U_t + g\pi_t \quad (\alpha, \beta > 0; 0 < g \leq 1) \quad (18.18)$$

Na equação de expectativas adaptativas, a derivada deve ser substituída por uma expressão de diferenças:

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \quad (0 < j \leq 1) \quad (18.19)$$

Pelo mesmo critério, a equação da política monetária deve ser mudada para[†]

$$U_{t+1} - U_t = -k(m - p_{t+1}) \quad (k > 0) \quad (18.20)$$

Essas três equações constituem a nova versão do modelo inflação-desemprego.

A equação de diferenças em p

Na primeira etapa da análise desse novo modelo, novamente tentamos condensar o modelo em uma única equação com uma única variável. Seja p essa variável. Desta maneira, concentraremos nossa atenção em (18.18). Entretanto, visto que (18.18) – ao contrário das duas outras equações – não descreve um padrão de variação por si mesma, cabe a nós criar tal padrão, o que fazemos calculando a diferença de p_t , isto é, tomando a primeira diferença de p_t segundo a definição

$$\Delta p_t \equiv p_{t+1} - p_t$$

Há duas etapas envolvidas nessa operação. A primeira é adiantar os índices de tempo em (18.18) em um período para obter

[†] Supusemos que a variação em U_t depende de $(m - p_{t+1})$, a taxa de crescimento da moeda real no período $(t + 1)$. Como alternativa, é possível fazer com que ela dependa da taxa de crescimento da moeda real no período t , $(m - p_t)$ (veja Exercício 18.3-4).

$$p_{t+1} = \alpha - T - \beta U_{t+1} + g\pi_{t+1} \quad (18.18')$$

Então subtraímos (18.18) de (18.18') para obter a diferença primeira de p , que dá o padrão de variação desejado:

$$\begin{aligned} p_{t+1} - p_t &= -\beta(U_{t+1} - U_t) + g(\pi_{t+1} - \pi_t) \\ &= \beta k(m - p_{t+1}) + gj(p_t - \pi_t) \quad [\text{por (18.20) e (18.19)}] \end{aligned} \quad (18.21)$$

Note que, na segunda linha de (18.21), os padrões de variação das outras duas variáveis, como dados em (18.19) e (18.20), foram incorporados ao padrão de variação da variável p . Assim, (18.21) agora incorpora todas as informações no presente modelo.

Todavia, o termo π_t é irrelevante para o estudo de p e precisa ser eliminado de (18.21). Com essa finalidade, empregamos o fato de que

$$g\pi_t = p_t - (\alpha - T) + \beta U_t \quad [\text{por (18.18)}] \quad (18.22)$$

Substituindo essa expressão em (18.21) e agrupando termos, obtemos

$$(1 + \beta k) p_{t+1} - [1 - j(1 - g)] p_t + j\beta U_t = \beta km + j(\alpha - T) \quad (18.23)$$

Mas agora aparece um termo U_t que tem de ser eliminado. Para fazer isso, calculamos a diferença de (18.23) para obter um termo $(U_{t+1} - U_t)$ e então usamos (18.20) para eliminar este último. Somente após esse processo razoavelmente longo de substituições é que obtemos a desejada equação de diferenças apenas na variável p . Essa equação, quando devidamente normalizada, assume a forma

$$p_{t+2} - \underbrace{\frac{1 + gj + (1 - j)(1 + \beta k)}{1 + \beta k}}_{a_1} p_{t+1} + \underbrace{\frac{1 - j(1 - g)}{1 + \beta k}}_{a_2} p_t = \underbrace{\frac{j\beta km}{1 + \beta k}}_c \quad (18.24)$$

A trajetória temporal de p

O valor de equilíbrio intertemporal de p , dado pela solução particular de (18.24), é

$$\bar{p} = \frac{c}{1 + a_1 + a_2} = \frac{j\beta km}{\beta kj} = m \quad [\text{por (18.2)}]$$

Portanto, assim como no modelo de tempo contínuo, a taxa de equilíbrio da inflação é exatamente igual à taxa de expansão monetária.

Quanto à função complementar, podem surgir ou raízes reais distintas (Caso 1), ou raízes reais repetidas (Caso 2), ou raízes complexas (Caso 3), dependendo das grandezas relativas de a_1^2 e $4a_2$. No presente modelo,

$$\begin{aligned} a_1^2 &\gtrless 4a_2 \quad \text{se, e somente se, } [1 + gj + (1 - j)(1 + \beta k)]^2 \\ &\gtrless 4[1 - j(1 - g)](1 + \beta k) \end{aligned} \quad (18.25)$$

Se $g = \frac{1}{2}$, $j = \frac{1}{3}$ e $\beta k = 5$, por exemplo, então $a_1^2 = (5\frac{1}{6})^2$, ao passo que $4a_2 = 20$; assim, resulta o Caso 1. Mas, se $g = j = 1$, então $a_1^2 = 4$ enquanto $4a_2 = 4(1 + \beta k) > 4$, e temos o Caso 3. Contudo, em vista do maior número de parâmetros no presente modelo, não é viável construir um gráfico classificatório como o da Figura 18.2 no modelo de Samuelson.

Não obstante, a análise de convergência ainda pode prosseguir segundo a mesma linha da Seção 18.2. Especificamente, recordamos, por (16.6), que as duas raízes características b_1 e b_2 devem satisfazer as duas relações seguintes:

$$b_1 + b_2 = -a_1 = \frac{1 + gj}{1 + \beta k} + 1 - j > 0 \quad (18.26)$$

[veja (18.24)]

$$b_1 b_2 = a_2 = \frac{1 - j(1 - g)}{1 + \beta k} \in (0, 1) \quad (18.26')$$

Além do mais, no presente modelo temos

$$(1 - b_1)(1 - b_2) = 1 - (b_1 + b_2) + b_1 b_2 = \frac{\beta jk}{1 + \beta k} > 0 \quad (18.27)$$

Agora considere o Caso 1, no qual as duas raízes b_1 e b_2 são reais e distintas. Uma vez que seu produto $b_1 b_2$ é positivo, b_1 e b_2 devem ter o mesmo sinal. Além do mais, como sua soma é positiva, b_1 e b_2 devem ser ambas positivas, o que implica que nenhuma oscilação pode ocorrer. Podemos inferir, por (18.27), que nem b_1 nem b_2 pode ser igual a um; caso contrário, $(1 - b_1)(1 - b_2)$ seria zero, o que violaria a desigualdade indicada. Isso significa que, em termos das várias possibilidades de combinações de (b_1, b_2) enumeradas no modelo de Samuelson, as Possibilidades *ii* e *iv* não podem surgir aqui. Também é inaceitável que uma das raízes seja maior que um e a outra menor que um; pois, caso contrário, $(1 - b_1)(1 - b_2)$ seria negativo. Assim, a Possibilidade *iii* também está descartada. Segue-se que b_1 e b_2 devem ser ambas maiores que um ou ambas menores que um. Contudo, se $b_1 > 1$ e $b_2 > 1$ (Possibilidade *v*), (18.26') seria violada. Por conseguinte, a única eventualidade viável é a Possibilidade *i*, isto é, b_1 e b_2 são ambos positivos e menores que 1, portanto a trajetória temporal de p é convergente.

A análise do Caso 2 não é basicamente muito diferente. Por raciocínio praticamente idêntico, podemos concluir que a raiz repetida b só pode ser positiva e menor que 1 nesse modelo; isto é, a Possibilidade *vi* é viável, mas as Possibilidades *vii* e *viii* não são. A trajetória temporal de p no Caso 2 novamente é não-oscilatória e convergente.

Para o Caso 3, a convergência requer que R (o valor absoluto das raízes complexas) seja menor que um. Por (18.8), $R = \sqrt{a_2}$. Considerando que a_2 é positivo e menor que 1 [veja (18.26')], realmente temos $R < 1$. Assim, a trajetória temporal de p no Caso 3 também é convergente, embora dessa vez haja flutuação de grau.

A análise de U

Se quisermos analisar a trajetória temporal da taxa de desemprego, podemos tomar (18.20) como ponto de partida. Para nos livrarmos do termo p daquela equação, primeiro substituímos (18.18') para obter

$$(1 + \beta k) U_{t+1} - U_t = k(\alpha - T - m) + kg\pi_{t+1} \quad (18.28)$$

Em seguida, para preparar a substituição da outra equação, (18.19), calculamos a diferença de (18.28) e constatamos que

$$(1 + \beta k) U_{t+2} - (2 + \beta k) U_{t+1} + U_t = kg(\pi_{t+2} - \pi_{t+1}) \quad (18.29)$$

Em vista da presença de uma expressão de diferença em π no lado direito, podemos substituí-la por uma versão deslocada para diante da equação de expectativas adaptativas. Isso resulta em

$$(1 + \beta k) U_{t+2} - (2 + \beta k) U_{t+1} + U_t = kgj(p_{t+1} - \pi_{t+1}) \quad (18.30)$$

que é a incorporação de todas as informações no modelo.

Contudo, temos de eliminar as variáveis p e π para que surja uma equação de diferenças adequada na variável U . Com essa finalidade, notamos, por (18.20), que

$$kp_{t+1} = U_{t+1} - U_t + km \quad (18.31)$$

Além disso, multiplicando todos os termos de (18.22) por $(-kj)$ e deslocando os índices de tempo, podemos escrever

$$\begin{aligned} -kjg\pi_{t+1} &= -kj p_{t+1} + kj(\alpha - T) - \beta k j U_{t+1} \\ &= -j(U_{t+1} - U_t + km) + kj(\alpha - T) - \beta k j U_{t+1} \\ &\quad [\text{por (18.31)}] \\ &= -j(1 + \beta k)U_{t+1} + jU_t + kj(\alpha - T - m) \end{aligned} \quad (18.32)$$

Esses dois resultados expressam p_{t+1} e π_{t+1} em termos da variável U e, assim, por substituição em (18.30), podem nos habilitar a obter – finalmente! – a equação de diferenças desejada na variável U apenas:

$$\begin{aligned} U_{t+2} - \frac{1 + gj + (1 - j)(1 + \beta k)}{1 + \beta k} U_{t+1} + \frac{1 - j(1 - g)}{1 + \beta k} U_t \\ = \frac{kj[\alpha - T - (1 - g)m]}{1 + \beta k} \end{aligned} \quad (18.33)$$

É digno de nota que os dois coeficientes constantes do lado esquerdo (a_1 e a_2) são idênticos aos da equação de diferenças para p [isto é, (18.24)]. O resultado é que a análise anterior da função complementar da trajetória de p deve ser igualmente aplicável ao presente contexto. Mas o termo constante do lado direito de (18.33) é diferente do de (18.24). Por consequência, as soluções particulares nas duas situações serão diferentes. E é assim que deve ser pois, à parte a coincidência, não há nenhuma razão inerente para esperar que a taxa de equilíbrio intertemporal do desemprego seja a mesma que a taxa de equilíbrio da inflação.

A relação de Phillips de longo prazo

Pode-se verificar, de pronto, que a taxa de equilíbrio intertemporal do desemprego é

$$\bar{U} = \frac{1}{\beta} [\alpha - T - (1 - g)m]$$

Mas, uma vez que constatamos que a taxa de equilíbrio da inflação é $\bar{p} = m$, podemos ligar $\bar{U}$ a $\bar{p}$ pela equação

$$\bar{U} = \frac{1}{\beta} [\alpha - T - (1 - g)\bar{p}] \quad (18.34)$$

Como essa equação diz respeito apenas às taxas de *equilíbrio* de desemprego e inflação, diz-se que ela representa a relação de Phillips de *longo prazo*.

Um caso especial de (18.34) vem sendo alvo de muita atenção entre economistas: o caso de $g = 1$. Se $g = 1$, o termo $\bar{p}$ terá um coeficiente zero e, desse modo, sairá do quadro. Em outras palavras, $\bar{U}$ se tornará uma função constante de $\bar{p}$. No diagrama padrão de Phillips, no qual a taxa de desemprego é colocada no eixo horizontal, esse resultado dá origem a uma curva de Phillips de longo prazo vertical. Nesse caso, o valor de $\bar{U}$, denominado *taxa natural de desemprego*, então é consistente com qualquer taxa de equilíbrio da inflação, além da notável implicação política de que, no longo prazo, não há nenhuma permuta entre os dois males, o da inflação e o do desemprego, tal como existe no curto prazo.

Mas, e se $g < 1$? Nesse caso, o coeficiente de $\bar{p}$ em (18.34) será negativo. Então, a curva de Phillips de longo prazo revelará ter inclinação descendente e, por conseguinte, ainda proporciona uma permuta entre inflação e desemprego. Portanto, o fato de a curva de Phillips de longo prazo ser vertical ou ter inclinação negativa depende criticamente do valor do parâmetro g , o qual, segundo a relação de Phillips com expectativas aumentadas, mede até que ponto a taxa de inflação esperada pode se infiltrar na estrutura salarial e na própria taxa de inflação real. Tudo isso pode lhe parecer muito familiar – isso porque discutimos o tópico no Exemplo 1 na Seção 16.5 e você também trabalhou com ele no Exercício 16.5-4.

EXERCÍCIO 18.3

1. Apresente as etapas intermediárias que levam de (18.23) a (18.24).
2. Mostre que, se o modelo discutido nesta seção for condensado em uma equação de diferenças na variável π , o resultado será o mesmo que em (18.24) exceto pela substituição de π por p .
3. Constatamos que as trajetórias temporais de p e U no modelo discutido nesta seção eram consistentemente convergentes. Podem surgir trajetórias temporais divergentes se descartarmos a premissa de que $g \leq 1$? Se a resposta for positiva, quais das “possibilidades” de divergência nos Casos 1, 2 e 3 serão viáveis agora?
4. Mantenha as equações (18.18) e (18.19), mas troque (18.20) para $U_{t+1} - U_t = -k(m - p_t)$
 - (a) Deduza uma nova equação de diferenças na variável p .
 - (b) Essa nova equação de diferenças dá um $\bar{p}$ diferente?
 - (c) Suponha que $j = g = 1$. Encontre as condições sob as quais as raízes características cairão nos Casos 1, 2 e 3, respectivamente.
 - (d) Seja $j = g = 1$. Descreva a trajetória temporal de p (incluindo convergência ou divergência) quando $\beta k = 3, 4$ e 5 , respectivamente.

18.4 Generalizações para equações de termo variável e ordens mais altas

Agora estamos prontos para estender nossos métodos em duas direções – o caso do termo variável e o caso das equações de diferenças de ordens mais altas.

Termo variável na forma de cm^t

Quando o termo constante c em (18.1) é substituído por um termo variável – alguma função de t – o único efeito será sobre a solução particular. (Por quê?) Para achar a nova solução particular, podemos aplicar novamente o método de coeficientes indeterminados. No contexto da equação diferencial (Seção 16.6), esse método requer que o termo variável e suas derivadas sucessivas assumam, em conjunto, somente um número finito de tipos distintos de expressão, à parte as constantes multiplicativas. Esse requisito, quando aplicado a equações de diferenças, deve ser corrigido para “o termo variável e suas *diferenças* sucessivas devem assumir, em conjunto, somente um número finito de tipos de expressão, à parte as constantes multiplicativas”. Vamos ilustrar esse método com exemplos concretos tomando, em primeiro lugar, um termo variável na forma cm^t , onde c e m são constantes.

EXEMPLO 1 Encontre a solução particular de

$$y_{t+2} + y_{t+1} - 3y_t = 7^t$$

Aqui, temos $c = 1$ e $m = 7$. Primeiro, vamos averiguar se o termo variável 7^t resulta em um número finito de tipos de expressões quando se calculam diferenças sucessivas. Segundo a regra da diferença ($\Delta y_t = y_{t+1} - y_t$), a diferença *primeira* do termo é

$$\Delta 7^t = 7^{t+1} - 7^t = (7 - 1)7^t = 6(7)^t$$

De modo semelhante, a diferença *segunda*, $\Delta^2(7^t)$, pode ser expressa como

$$\Delta(\Delta 7^t) = \Delta 6(7^t) = 6(7)^{t+1} - 6(7)^t = 6(7 - 1)7^t = 36(7)^t$$

Além disso, como pode ser verificado, todas as diferenças sucessivas, assim como a primeira e a segunda, serão algum múltiplo de 7^t . Uma vez que há somente um único tipo de expressão, podemos tentar uma solução $y_t = B(7)^t$ para a solução particular, onde B é um coeficiente indeterminado.

Substituindo a solução experimental e suas versões correspondentes para os períodos $(t + 1)$ e $(t + 2)$ na equação de diferenças dada, obtemos

$$B(7)^{t+2} + B(7)^{t+1} - 3B(7)^t = 7^t \quad \text{ou} \quad B(7^2 + 7 - 3)(7)^t = 7^t$$

Assim,

$$B = \frac{1}{49 + 7 - 3} = \frac{1}{53}$$

e podemos escrever a solução particular como

$$y_p = B(7)^t = \frac{1}{53}(7)^t$$

Essa expressão pode ser tomada como um equilíbrio móvel. Você pode verificar a correção da solução substituindo-a na equação de diferenças e averiguando se isso resultará em uma identidade $7^t = 7^t$.

O resultado que encontramos no Exemplo 1 pode ser generalizado com facilidade do termo variável 7^t para o termo cm^t . Esperamos, por nossa experiência, que todas as diferenças sucessivas de cm^t assumam a mesma forma de expressão, a saber, Bm^t , onde B é alguma constante multiplicativa. Por conseguinte, podemos tentar uma solução $y_t = Bm^t$ para a solução particular, dada a equação de diferenças

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = cm^t \quad (18.35)$$

Usando a solução experimental $y_t = Bm^t$, o que implica $y_{t+1} = Bm^{t+1}$ etc., podemos reescrever a equação (18.35) como

$$Bm^{t+2} + a_1 Bm^{t+1} + a_2 Bm^t = cm^t$$

ou

$$B(m^2 + a_1 m + a_2)m^t = cm^t$$

Por conseguinte, o coeficiente B na solução experimental deve ser

$$B = \frac{c}{m^2 + a_1 m + a_2}$$

e a solução particular desejada de (18.35) pode ser escrita como

$$y_p = Bm^t = \frac{c}{m^2 + a_1 m + a_2} m^t \quad (m^2 + a_1 m + a_2 \neq 0) \quad (18.36)$$

Note que não é permitido que o denominador de B seja zero. Caso isso aconteça,[†] então devemos usar a solução experimental $y^t = Btm^t$; ou, se esta também falhar, $y^t = Bt^2m^t$.

Termo variável na forma de ct^n

Agora vamos considerar termos variáveis na forma ct^n , onde c é qualquer constante e n é um inteiro positivo.

EXEMPLO 2 Encontre a solução particular de

$$y_{t+2} + 5y_{t+1} + 2y_t = t^2$$

As primeiras três diferenças de t^2 (um caso especial de ct^n com $c = 1$ e $n = 2$) são calculados como se segue:[‡]

$$\begin{aligned}\Delta t^2 &= (t+1)^2 - t^2 = 2t + 1 \\ \Delta^2 t^2 &= \Delta(\Delta t^2) = \Delta(2t + 1) = \Delta 2t + \Delta 1 \\ &= 2(t+1) - 2t + 0 = 2 \quad [\Delta \text{ constante} = 0] \\ \Delta^3 t^2 &= \Delta(\Delta^2 t^2) = \Delta 2 = 0\end{aligned}$$

Uma vez que continuar calculando diferenças somente resultará zero, há, no total, três tipos distintos de expressão: t^2 (do próprio termo variável), t e uma constante (pelas diferenças sucessivas). Portanto, vamos tentar a solução

$$y_t = B_0 + B_1 t + B_2 t^2$$

para a solução particular, com coeficientes indeterminados B_0 , B_1 e B_2 . Note que essa solução implica

$$\begin{aligned}y_{t+1} &= B_0 + B_1(t+1) + B_2(t+1)^2 \\ &= (B_0 + B_1 + B_2) + (B_1 + 2B_2)t + B_2 t^2 \\ y_{t+2} &= B_0 + B_1(t+2) + B_2(t+2)^2 \\ &= (B_0 + 2B_1 + 4B_2) + (B_1 + 4B_2)t + B_2 t^2\end{aligned}$$

Quando essas expressões são substituídas na equação de diferenças, obtemos

$$(8B_0 + 7B_1 + 9B_2) + (8B_1 + 14B_2)t + 8B_2 t^2 = t^2$$

Igualando os dois lados termo a termo, vemos que os coeficientes indeterminados têm de satisfazer as seguintes equações simultâneas:

$$\begin{aligned}8B_0 + 7B_1 + 9B_2 &= 0 \\ 8B_1 + 14B_2 &= 0 \\ 8B_2 &= 1\end{aligned}$$

[†] Essa situação, análoga à do Exemplo 3 da Seção 16.6, se materializará eventualmente, quando acontecer de a constante m ser igual à raiz característica da equação de diferenças. As raízes características da equação de diferenças de (18.35) são os valores de b que satisfazem a equação $b^2 + a_1 b + a_2 = 0$. Se acaso uma das raízes tiver o valor m , então devemos deduzir que $m^2 + a_1 m + a_2 = 0$.

[‡] Esses resultados devem ser comparados com as três primeiras derivadas de t^2 :

$$\frac{d}{dt} t^2 = 2t \quad \frac{d^2}{dt^2} t^2 = 2 \quad \text{e} \quad \frac{d^3}{dt^3} t^2 = 0$$

Assim, seus valores devem ser $B_0 = \frac{13}{256}$, $B_1 = -\frac{7}{32}$ e $B_2 = \frac{1}{8}$, o que nos dá a solução particular

$$y_p = \frac{13}{256} - \frac{7}{32}t + \frac{1}{8}t^2$$

Nossa experiência com o termo variável t^2 deve nos habilitar a generalizar o método para o caso de ct^n . Na nova solução experimental, é óbvio que deve haver um termo B_nt^n para corresponder ao termo variável dado. Além do mais, visto que o cálculo de diferenças sucessivas do termo dá como resultado as expressões distintas t^{n-1} , t^{n-2} , ..., t e B_0 (constante), a nova solução experimental para o caso do termo variável ct^n deve ser escrita como

$$y_t = B_0 + B_1t + B_2t^2 + \cdots + B_nt^n$$

Mas o restante do procedimento é inteiramente o mesmo.

É preciso acrescentar que tal solução experimental também pode falhar. Nesse caso, o truque – já empregado em inúmeras outras ocasiões – é novamente multiplicar a solução experimental original por uma potência suficientemente alta de t . Isto é, podemos tentar, por sua vez, $y_t = t(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)$ etc.

Equações de diferenças lineares de ordens mais altas

A *ordem* de uma equação de diferenças indica a diferença de ordem mais alta presente na equação; mas também indica o número máximo de períodos de defasagem de tempo envolvidos. Assim, uma equação de diferenças linear de n -ésima ordem (com coeficientes constantes e termo constante) pode ser escrita, em geral, como

$$y_{t+n} + a_1 y_{t+n-1} + \cdots + a_{n-1} y_{t+1} + a_n y_t = c \quad (18.37)$$

O método para encontrar a solução particular dessa equação não é diferente de nenhum modo substantivo. Para começar, ainda podemos tentar $y_t = k$ (o caso de equilíbrio intertemporal constante). Se essa tentativa falhar, então experimentamos $y_t = kt$ ou $y_t = kt^2$ etc., nessa ordem.

Contudo, na busca da função complementar, agora vamos nos defrontar com uma equação característica que é uma equação polinomial de n -ésimo grau:

$$b^n + a_1 b^{n-1} + \cdots + a_{n-1} b + a_n = 0 \quad (18.38)$$

Agora haverá n raízes características b_i ($i = 1, 2, \dots, n$), todas as quais devem entrar na função complementar do seguinte modo:

$$y_c = \sum_{i=1}^n A_i b_i^t \quad (18.39)$$

contanto, é claro, que as raízes sejam todas reais e distintas. Caso haja raízes reais repetidas (digamos, $b_1 = b_2 = b_3$), então os três primeiros termos da soma em (18.39) devem ser modificados para

$$A_1 b_1^t + A_2 t b_1^t + A_3 t^2 b_1^t \quad [\text{cf. (18.6)}]$$

Além disso, se houver um par de raízes complexas conjugadas – digamos, b_{n-1} , b_n – então os dois últimos termos da soma em (18.39) devem ser combinados na expressão

$$R^t (A_{n-1} \cos \theta t + A_n \sin \theta t)$$

Uma expressão análoga também pode ser atribuída a qualquer outro par de raízes complexas. Caso haja dois pares *repetidos*, entretanto, um dos dois deve receber um fator multiplicativo, $t R^t$ em vez de R^t .

Após encontrar ambas, y_p e y_c , a solução geral da equação de diferenças completa (18.37) é novamente obtida por adição, isto é,

$$y_t = y_p + y_c$$

Mas, visto que haverá um total de n constantes arbitrárias nessa solução, serão exigidas nada menos do que n condições iniciais para defini-la.

EXEMPLO 3 Encontre a solução geral da equação de diferenças de terceira ordem

$$y_{t+3} - \frac{7}{8}y_{t+2} + \frac{1}{8}y_{t+1} + \frac{1}{32}y_t = 9$$

Experimentando a solução $y_t = k$, constatamos, com facilidade, que a solução particular é $y_p = 32$. Quanto à função complementar, uma vez que a equação característica cúbica

$$b^3 - \frac{7}{8}b^2 + \frac{1}{8}b + \frac{1}{32} = 0$$

pode ser fatorada para a forma

$$\left(b - \frac{1}{2}\right)\left(b - \frac{1}{2}\right)\left(b + \frac{1}{8}\right) = 0$$

as raízes são $b_1 = b_2 = \frac{1}{2}$ e $b_3 = -\frac{1}{8}$. Isso nos habilita a escrever

$$y_c = A_1\left(\frac{1}{2}\right)^t + A_2t\left(\frac{1}{2}\right)^t + A_3\left(-\frac{1}{8}\right)^t$$

Note que o segundo termo contém um t multiplicativo; isso se deve à presença de raízes repetidas. Então, a solução geral da equação de diferenças dada é simplesmente a soma de y_c e y_p .

Neste exemplo, acontece que todas as três raízes características são menores que 1 em valores absolutos. Por conseguinte, podemos concluir que a solução obtida representa uma trajetória temporal que converge para o nível de equilíbrio estacionário 32.

Convergência e o teorema de Schur

Quando temos uma equação de diferenças de ordem alta que não é resolvida com facilidade, ainda assim podemos determinar a convergência da trajetória temporal relativa em termos qualitativos, sem ter de enfrentar sua solução quantitativa propriamente dita. Lembre-se que a trajetória temporal pode convergir se, e somente se, cada uma das raízes da equação característica for menor que 1 em valor absoluto. Em vista disso, o seguinte teorema – conhecido como *teorema de Schur*[†] – torna-se diretamente aplicável:

As raízes da equação polinomial de n -ésimo grau

$$a_0b^n + a_1b^{n-1} + \dots + a_{n-1}b + a_n = 0$$

serão todas menores que a unidade em valor absoluto se, e somente se, os n determinantes seguintes

$$\Delta_1 = \begin{vmatrix} a_0 & a_n \\ a_n & a_0 \end{vmatrix} \quad \Delta_2 = \begin{vmatrix} a_0 & 0 & a_n & a_{n-1} \\ a_1 & a_0 & 0 & a_n \\ \dots & \dots & \dots & \dots \\ a_n & 0 & a_0 & a_1 \\ a_{n-1} & a_n & 0 & a_0 \end{vmatrix} \quad \dots$$

[†] Caso queira uma discussão desse teorema e sua história, consulte Chipman, John S. *The Theory of Inter-Sectoral Money Flows and Income Formation*. Baltimore: The Johns Hopkins Press, 1951, pp. 119–120.

$$\Delta_n = \begin{vmatrix} a_0 & 0 & \dots & 0 & a_n & a_{n-1} & \dots & a_1 \\ a_1 & a_0 & \dots & 0 & 0 & a_n & \dots & a_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n-1} & a_{n-2} & \dots & a_0 & 0 & 0 & \dots & a_n \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_n & 0 & \dots & 0 & a_0 & a_1 & \dots & a_{n-1} \\ a_{n-1} & a_n & \dots & 0 & 0 & a_0 & \dots & a_{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_1 & a_2 & \dots & a_n & 0 & 0 & \dots & a_0 \end{vmatrix}$$

forem todos positivos.

Note que, uma vez que a condição no teorema é expressa na base de “se e somente se”, ela é uma condição necessária e suficiente. Assim, o teorema de Schur é uma contraparte perfeita da equação de diferenças do teorema de Routh apresentado anteriormente na estrutura da equação diferencial.

A construção desses determinantes se baseia em um procedimento simples, que é melhor explicado com o auxílio das linhas pontilhadas que dividem cada determinante em quatro áreas. Cada área do k -ésimo determinante, Δ_k , sempre consiste em um subdeterminante $k \times k$. A área superior esquerda tem apenas a_0 na diagonal, zeros acima da diagonal, e índices progressivamente maiores para os coeficientes sucessivos em cada coluna abaixo dos elementos na diagonal. Quando transpomos os elementos da área superior esquerda, obtemos a área inferior direita. Voltando à área superior direita, agora colocamos o coeficiente a_n sozinho na diagonal, com zeros abaixo da diagonal e índices progressivamente menores para os coeficientes sucessivos à medida que subimos cada coluna a partir da diagonal. Quando os elementos dessa área são transpostos, obtemos a área inferior esquerda.

A aplicação desse teorema é direta. Uma vez que os coeficientes da equação característica são os mesmos que os coeficientes que aparecem no lado esquerdo da equação de diferenças original, podemos introduzi-los diretamente nos determinantes citados. Note que, em nosso contexto, sempre temos $a_0 = 1$.

EXEMPLO 4 A trajetória temporal da equação $y_{t+2} + 3y_{t+1} + 2y_t = 12$ converge? Aqui, temos $n = 2$, e os coeficientes são $a_0 = 1$, $a_1 = 3$ e $a_2 = 2$. Assim, obtemos

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3 < 0$$

Uma vez que isso já viola a condição de convergência, não há nenhuma necessidade de passar para Δ_2 .

Na verdade, podemos constatar com facilidade que as raízes características da equação de diferenças dada são $b_1, b_2 = -1, -2$, o que, de fato, implica uma trajetória temporal divergente.

EXEMPLO 5 Teste a convergência da trajetória de $y_{t+2} + \frac{1}{6}y_{t+1} - \frac{1}{6}y_t = 2$ pelo teorema de Schur. Aqui, os coeficientes são $a_0 = 1$, $a_1 = \frac{1}{6}$, $a_2 = -\frac{1}{6}$ (com $n = 2$). Assim, temos

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 \end{vmatrix} = \frac{35}{36} > 0$$

$$\Delta_2 = \begin{vmatrix} a_0 & 0 & a_2 & a_1 \\ a_1 & a_0 & 0 & a_2 \\ a_2 & 0 & a_0 & a_1 \\ a_1 & a_2 & 0 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 1 & 0 & -\frac{1}{6} \\ -\frac{1}{6} & 0 & 1 & \frac{1}{6} \\ \frac{1}{6} & -\frac{1}{6} & 0 & 1 \end{vmatrix} = \frac{1.176}{1.296} > 0$$

Esses resultados satisfazem a condição necessária e suficiente para convergência.

EXERCÍCIO 18.4

1. Aplique a definição do símbolo de “diferença” para achar:
 (a) Δt (b) $\Delta^2 t$ (c) Δt^3
 Compare os resultados da diferença com os da diferenciação.
2. Encontre a solução particular de cada uma das seguintes:
 (a) $y_{t+2} + 2y_{t+1} + y_t = 3^t$
 (b) $y_{t+2} - 5y_{t+1} - 6y_t = 2(6)^t$
 (c) $3y_{t+2} + 9y_t = 3(4)^t$
3. Encontre as soluções particulares de:
 (a) $y_{t+2} - 2y_{t+1} + 5y_t = t$
 (b) $y_{t+2} - 2y_{t+1} + 5y_t = 4 + 2t$
 (c) $y_{t+2} + 5y_{t+1} + 2y_t = 18 + 6t + 8t^2$
4. É de se esperar que, quando o termo variável toma a forma $m^t + t^n$, a solução experimental seja $B(m)^t + (B_0 + B_1 t + \dots + B_n t^n)$? Por quê?
5. Encontre as raízes características e a função complementar de:
 (a) $y_{t+3} - \frac{1}{2}y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 0$
 (b) $y_{t+3} - 2y_{t+2} + \frac{5}{4}y_{t+1} - \frac{1}{4}y_t = 1$
 [Sugestão: Tente fatorar $(b - \frac{1}{4})$ em ambas as equações características.]
6. Teste a convergência das soluções das seguintes equações de diferenças pelo teorema de Schur:
 (a) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 3$
 (b) $y_{t+2} - \frac{1}{9}y_t = 1$
7. No caso de uma equação de diferenças de terceira ordem
 $y_{t+3} + a_1 y_{t+2} + a_2 y_{t+1} + a_3 y_t = c$
 quais são as formas exatas dos determinantes requeridos pelo teorema de Schur?

CAPÍTULO 19

Equações diferenciais e equações de diferenças simultâneas

Até aqui, o que discutimos sobre a dinâmica da economia se restringiu à análise de uma equação dinâmica única (diferencial ou de diferenças). No presente capítulo, são apresentados métodos para analisar um sistema de equações dinâmicas simultâneas. Como isso acarretaria a manipulação de diversas variáveis ao mesmo tempo, é compreensível que você espere uma grande quantidade de novas complicações. Mas a verdade é que grande parte do que já aprendemos sobre equações dinâmicas únicas pode ser estendido de imediato a sistemas de equações dinâmicas simultâneas. Por exemplo, a solução de um sistema dinâmico consistiria em um conjunto de soluções particulares (valores de equilíbrio intertemporal das várias variáveis) e funções complementares (desvios em relação aos equilíbrios). As funções complementares ainda seriam baseadas nas equações associadas às homogêneas equações no sistema. E a estabilidade dinâmica do sistema ainda dependeria dos sinais (em sistemas de equações diferenciais) ou dos valores absolutos (em sistemas de equações de diferenças) das raízes características nas funções complementares. Assim, o problema de um sistema dinâmico é apenas ligeiramente mais complicado que o de uma equação dinâmica única.

19.1 A gênese de sistemas dinâmicos

Um sistema dinâmico pode surgir por dois modos gerais. Ele pode emanar de um dado *conjunto* de padrões de variação interativos ou ser derivado de um *único* padrão de variação dado, contanto que este último consista em uma equação dinâmica de segunda ordem (ou de ordem mais alta).

Padrões de variação interativos

O caso mais óbvio de um conjunto dado de padrões de variação interativos é o de um modelo multissetorial, no qual cada setor, como descrito por uma equação dinâmica, interfere em ao menos um dos outros setores. Uma versão dinâmica do modelo de insumo-produção, por exemplo, poderia envolver n indústrias cujas variações nos resultados produzem repercussões dinâmicas nas outras indústrias. Assim, ele constitui um sistema dinâmico. De modo semelhante, um modelo de mercado de equilíbrio dinâmico geral envolveria n mercadorias que se inter-relacionam por seus ajustes de preços. Assim, há novamente um sistema dinâmico.

Contudo, padrões de variação interativos podem ser encontrados até mesmo em um modelo de um único setor. As diversas variáveis em tal modelo não representam setores diferentes nem mercadorias diferentes, mas *aspectos* diferentes de uma economia. Ainda assim eles podem

afetar um ao outro em seu comportamento dinâmico, de modo a proporcionar uma rede de interações.[†] Na verdade, encontramos um exemplo concreto disso no Capítulo 18. No modelo de inflação-desemprego, a taxa de inflação esperada π segue um padrão de variação, (18.19), que depende não somente de π , mas também da taxa de desemprego U (por meio da taxa de inflação real p). Reciprocamente, o padrão de variação de U , (18.20), depende de π (novamente por meio de p). Assim, a dinâmica de π e U deve ser determinada simultaneamente. Em retrospectiva, portanto, o modelo de inflação-desemprego poderia ter sido tratado como um modelo dinâmico de equações simultâneas. E isso teria impedido a longa sequência de substituições e eliminações que foram executadas para condensar o modelo em uma única equação de uma só variável. De fato, logo adiante, na Seção 19.4, vamos retrabalhar aquele modelo por meio da perspectiva de um sistema dinâmico. Por enquanto, a noção de que o mesmo modelo pode ser analisado como uma equação única ou como um sistema de equações nos indica naturalmente como será a discussão do segundo modo de ter um sistema dinâmico.

A transformação de uma equação dinâmica de ordem alta

Suponha que temos uma equação diferencial (ou de diferenças) de n -ésima ordem em *uma* variável. Então, como demonstraremos em breve, é sempre possível transformar essa equação em um sistema matematicamente equivalente de n equações diferenciais (ou de diferenças) simultâneas de *primeira* ordem em n variáveis. Em particular, uma equação diferencial de segunda ordem pode ser reescrita como duas equações diferenciais de primeira ordem simultâneas em duas variáveis.[‡] Portanto, mesmo que aconteça de começarmos com somente uma equação dinâmica (de ordem alta), ainda assim é possível obter um sistema dinâmico por meio do artifício da transformação matemática. A propósito, esse fato tem uma importante implicação: na discussão de sistemas dinâmicos que se segue, basta que nos preocupemos com sistemas de equações de *primeira ordem*, pois, se estiver presente uma equação de ordem mais alta, sempre podemos transformá-la, antes de mais nada, em um conjunto de equações de primeira ordem. Isso resultará em um número maior de equações no sistema, mas, por outro lado, a ordem será reduzida ao mínimo.

Para ilustrar o procedimento de transformação, vamos considerar a equação de diferenças única

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \quad (19.1)$$

Se inventarmos uma nova variável artificial x_t , definida por

$$x_t \equiv y_{t+1} \quad (\text{implicando } x_{t+1} \equiv y_{t+2})$$

então podemos expressar a equação de segunda ordem original por meio de *duas* equações de primeira ordem simultâneas (defasagem de um período) da seguinte maneira:

$$\begin{array}{rcl} x_{t+1} & + & a_1 x_t + a_2 y_t = c \\ y_{t+1} & - & x_t = 0 \end{array} \quad (19.1')$$

É fácil ver que, desde que a segunda equação (que define a variável x_t) seja satisfeita, a primeira é idêntica à equação original dada. Por um procedimento análogo, e utilizando mais variáveis artificiais, podemos transformar, de modo semelhante, uma equação única de ordem mais alta em um sistema equivalente de equações simultâneas de primeira ordem. Você pode verificar, por exemplo, que a equação de terceira ordem

$$y_{t+3} + y_{t+2} - 3y_{t+1} + 2y_t = 0 \quad (19.2)$$

[†] Note que, se tivermos duas equações dinâmicas nas duas variáveis y_1 e y_2 tais que o padrão de variação de y_1 dependa exclusivamente do próprio y_1 , o mesmo ocorrendo com y_2 , na verdade não temos um sistema de equações simultâneas. Em vez disso, temos apenas duas meras equações dinâmicas isoladas, cada uma das quais pode ser analisada por si, sem nenhum requisito de "simultaneidade".

[‡] Reciprocamente, duas equações diferenciais (ou de diferenças) de primeira ordem em duas variáveis podem ser consolidadas em uma única equação de segunda ordem de uma só variável, como fizemos nas Seções 16.5 e 18.3.

pode ser expressa como

$$\begin{aligned} w_{t+1} + w_t - 3x_t + 2y_t &= 0 \\ x_{t+1} - w_t &= 0 \\ y_{t+1} - x_t &= 0 \end{aligned} \quad (19.2')$$

onde $x_t \equiv y_{t+1}$ (de modo que $x_{t+1} \equiv y_{t+2}$) e $w_t \equiv x_{t+1}$ (de modo que $w_{t+1} \equiv x_{t+2} \equiv y_{t+3}$).

Por um procedimento perfeitamente análogo, também podemos transformar uma equação diferencial de n -ésima ordem em um sistema de n equações de primeira ordem. Dada a equação diferencial de segunda ordem

$$y''(t) + a_1 y'(t) + a_2 y(t) = 0 \quad (19.3)$$

por exemplo, podemos introduzir uma nova variável $x(t)$, definida por

$$x(t) \equiv y'(t) \quad [\text{implicando } x'(t) \equiv y''(t)]$$

Então, (19.3) pode ser reescrita como o seguinte sistema de duas equações de primeira ordem:

$$\begin{aligned} x'(t) + a_1 x(t) + a_2 y(t) &= 0 \\ y'(t) - x(t) &= 0 \end{aligned} \quad (19.3')$$

no qual, você pode notar, a segunda equação desempenha a função de definir a variável x recém-introduzida, como o fez a segunda equação em (19.1'). Em essência, o mesmo procedimento também pode ser utilizado para transformar uma equação diferencial de ordem mais alta. A única modificação é que temos de introduzir um número maior correspondente de novas variáveis.

19.2 Resolvendo equações dinâmicas simultâneas

Os métodos para resolver equações diferenciais simultâneas e equações de diferenças simultâneas são bastante semelhantes e, por isso, vamos discuti-los em conjunto nesta seção. Para nossas finalidades presentes, limitaremos a discussão a equações lineares com coeficientes constantes, apenas.

Equações de diferenças simultâneas

Suponha que temos o seguinte sistema de equações de diferenças lineares:

$$\begin{aligned} x_{t+1} + 6x_t + 9y_t &= 4 \\ y_{t+1} - x_t &= 0 \end{aligned} \quad (19.4)$$

Como obter as trajetórias temporais de x e y tais que ambas as equações nesse sistema sejam satisfeitas? Em essência, nossa tarefa é, mais uma vez, procurar as soluções particulares e funções complementares e somá-las para obter as trajetórias temporais desejadas das duas variáveis.

Uma vez que soluções particulares representam valores de equilíbrio intertemporal, vamos denotá-las por $\bar{x}$ e $\bar{y}$. Como antes, é aconselhável tentar primeiro soluções constantes, a saber, $x_{t+1} = x_t = \bar{x}$ e $y_{t+1} = y_t = \bar{y}$. E isso realmente funcionará no presente caso pois, substituindo essas soluções experimentais em (19.4), obtemos

$$\begin{cases} 7\bar{x} + 9\bar{y} = 4 \\ -\bar{x} + \bar{y} = 0 \end{cases} \Rightarrow \bar{x} = \bar{y} = \frac{1}{4} \quad (19.5)$$

(Entretanto, caso essas soluções constantes não funcionem, então devemos tentar soluções da forma $x_t = k_1 t$, $y_t = k_2 t$ etc.)

Para as funções complementares, nossa experiência anterior nos diz que devemos adotar soluções experimentais da forma

$$x_t = mb^t \quad \text{e} \quad y_t = nb^t \quad (19.6)$$

onde m e n são constantes arbitrárias e a base b representa a raiz característica. Então, fica automaticamente subentendido que

$$x_{t+1} = mb^{t+1} \quad \text{e} \quad y_{t+1} = nb^{t+1} \quad (19.7)$$

Note que, para simplificar as coisas, estamos empregando a mesma base $b \neq 0$ para ambas as variáveis, embora sejam permitidos coeficientes diferentes. Nossa meta é achar os valores de b , m e n que possam fazer com que as soluções experimentais (19.6) satisfaçam a versão *homogênea* de (19.4).

Substituindo as soluções experimentais na versão homogênea de (19.4) e cancelando o fator comum $b^t \neq 0$, obtemos as duas equações

$$\begin{aligned} (b+6)m + 9n &= 0 \\ -m + bn &= 0 \end{aligned} \quad (19.8)$$

Essas expressões podem ser consideradas como um sistema de equações homogêneas lineares nas duas variáveis m e n – se nos dispusermos a considerar b como um parâmetro por enquanto. Como o sistema (19.8) é homogêneo, ele pode dar como resultado somente a solução trivial $m = n = 0$ se sua matriz de coeficientes for invertível (veja Tabela 5.1 na Seção 5.5). Nesse caso, as funções complementares em (19.6) serão ambas identicamente zero, significando que x e y nunca se desviam de seus valores de equilíbrio intertemporal. Uma vez que esse seria um caso especial desinteressante, tentaremos descartar aquela solução trivial exigindo que a matriz de coeficientes do sistema seja *singular*. Isto é, exigiremos que o determinante daquela matriz se anule:

$$\begin{vmatrix} b+6 & 9 \\ -1 & b \end{vmatrix} = b^2 + 6b + 9 = 0 \quad (19.9)$$

Por essa equação quadrática, constatamos que $b (= b_1 = b_2) = -3$ é o único valor que pode impedir que m e n sejam ambas zero em (19.8). Por conseguinte, usaremos somente esse valor de b . A equação (19.9) é denominada a *equação característica*, e suas raízes são denominadas as *raízes características* do sistema de equações de diferenças simultâneas dado.

Tão logo tenhamos um valor específico de b , (19.8) nos dará os valores de solução correspondentes de m e n . Contudo, como o sistema é homogêneo, na verdade surgirá um número infinito de soluções para (m, n) , que pode ser expresso na forma de uma equação $m = kn$, onde k é uma constante. De fato, para cada raiz b_i , em geral haverá uma equação distinta $m_i = k_i n_i$. Mesmo com raízes repetidas, ou seja, com $b_1 = b_2$, ainda assim usaríamos duas equações desse tipo, $m_1 = k_1 n_1$ e $m_2 = k_2 n_2$ nas funções complementares. Além do mais, no caso de raízes repetidas, lembramos, por (18.6), que as funções complementares devem ser escritas como

$$\begin{aligned} x_t &= m_1(-3)^t + m_2 t(-3)^t \\ y_t &= n_1(-3)^t + n_2 t(-3)^t \end{aligned}$$

Os fatores de proporcionalidade entre m_i e n_i devem, é claro, satisfazer o sistema de equações dado (19.4), que determina que $y_{t+1} = x_t$, isto é,

$$n_1(-3)^{t+1} + n_2(t+1)(-3)^{t+1} - m_1(-3)^t + m_2 t(-3)^t$$

Dividindo tudo por $(-3)^t$, obtemos

$$-3n_1 - 3n_2(t+1) = m_1 + m_2t$$

ou, após rearranjar,

$$-3(n_1 + n_2) - 3n_2t - m_1 + m_2t$$

Igualando os termos com t dos dois lados do sinal de igual, e, de forma semelhante, os termos sem t , encontramos

$$m_1 = -3(n_1 + n_2) \quad \text{e} \quad m_2 = -3n_2$$

Se agora escrevermos $n_1 = A_3$, $n_2 = A_4$, então segue-se que

$$m_1 = -3(A_3 + A_4) \quad m_2 = -3A_4$$

Assim, as funções complementares podem ser escritas como

$$\begin{aligned} x_c &= -3(A_3 + A_4)(-3)^t - 3A_4t(-3)^t \\ &= -3A_3(-3)^t - 3A_4(t+1)(-3)^t \\ y_c &= A_3(-3)^t + A_4t(-3)^t \end{aligned} \quad (19.10)$$

onde A_3 e A_4 são constantes arbitrárias. Então, a solução geral é deduzida com facilidade combinando as soluções particulares em (19.5) com as funções complementares que acabamos de encontrar. Resta apenas definir as duas constantes arbitrárias A_3 e A_4 com a ajuda de condições iniciais ou de fronteira adequadas.

Um aspecto significativo da solução precedente é que, uma vez que ambas as trajetórias temporais contêm expressões idênticas b^t , ambas devem convergir ou ambas devem divergir. Isso faz sentido porque, em um modelo com variáveis dinamicamente interdependentes, um equilíbrio intertemporal dinâmico não pode prevalecer a menos que nenhum movimento dinâmico esteja presente em nenhum lugar do sistema. No presente caso, com raízes repetidas $b = -3$, as trajetórias temporais de ambas, x e y , apresentarão oscilação explosiva.

Notação matricial

Para destacar o paralelismo básico entre os métodos de resolução de uma equação única e de um sistema de equações, a exposição precedente foi realizada sem o benefício de notação matricial. Agora vamos ver como essa notação pode ser utilizada aqui. Ainda que pareça não ter sentido aplicar notação matricial a um sistema simples de apenas duas equações, a possibilidade de estender essa notação para o caso de n equações deve ser suficiente para fazer com que o exercício valha a pena.

Antes de mais nada, o sistema dado (19.4) pode ser expresso como

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} + \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \quad (19.4')$$

ou, mais sucintamente, como

$$Iu + Kv = d \quad (19.4'')$$

onde I é a matriz identidade 2×2 ; K é a matriz 2×2 dos coeficientes dos termos x_t e y_t e u , v e d são vetores coluna definidos da seguinte maneira:[†]

$$u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} \quad v = \begin{bmatrix} x_t \\ y_t \end{bmatrix} \quad d = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

O leitor talvez ache uma das características intrigante: já que $Iu = u$, porque não descartar o P ? A resposta é que, embora pareça redundante agora, a matriz identidade será necessária em operações subseqüentes e, portanto, vamos conservá-la como em (19.4'').

Quando tentamos soluções constantes $x_{t+1} = x_t = \bar{x}$ e $y_{t+1} = y_t = \bar{y}$ para as soluções particulares, na verdade estamos estabelecendo $u = v = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$, o que reduzirá (19.4'') a

$$(I + K) \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = d$$

Se a inversa $(I + K)^{-1}$ existir, podemos expressar as soluções particulares como

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = (I + K)^{-1} d \quad (19.5')$$

É claro que essa é uma fórmula geral, pois é válida para qualquer matriz K e vetor d contanto que $(I + K)^{-1}$ exista. Aplicada a nosso exemplo numérico, temos

$$(I + K)^{-1} d = \begin{bmatrix} 7 & 9 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & -\frac{9}{16} \\ \frac{1}{16} & \frac{7}{16} \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

Portanto, $\bar{x} = \bar{y} = \frac{1}{4}$, o que está de acordo com (19.5).

Quanto às funções complementares, vemos que as soluções experimentais (19.6) e (19.7) dão aos vetores u e v as formas específicas

$$u = \begin{bmatrix} mb^{t+1} \\ nb^{t+1} \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} \quad \text{e} \quad v = \begin{bmatrix} mb^t \\ nb^t \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^t$$

Quando substituídas na equação homogênea $Iu + Kv = 0$, essas soluções experimentais transformarão a última em

$$I \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} + K \begin{bmatrix} m \\ n \end{bmatrix} b^t = 0$$

ou, após multiplicar tudo por b^{-t} (um escalar) e fatorar,

$$(bI + K) \begin{bmatrix} m \\ n \end{bmatrix} = 0 \quad (19.8')$$

onde 0 é um vetor zero. É por esse sistema de equações homogêneas que vamos encontrar os valores apropriados de b , m e n que deverão ser usados nas soluções experimentais de modo a fazer delas soluções determinadas.

[†] Aqui, o símbolo v denota um vetor. Não o confunda com o v na notação de números complexos $h \pm vj$, onde ele representa um escalar.

Para evitar soluções triviais para m e n , é necessário que

$$|bI + K| = 0 \quad (19.9')$$

E essa é a equação característica que nos dará as raízes características b_i . Você pode verificar que, se substituirmos

$$bI = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} \quad \text{e} \quad K = \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix}$$

nessa equação, o resultado será exatamente (19.9), dando as raízes repetidas $b = -3$.

Em geral, cada raiz b_i extrairá de (19.8') um conjunto particular de número infinito de valores de solução de m e n que estão ligados entre si pela equação $m_i = k_i n_i$. Por conseguinte, é possível escrever, para cada valor de b_i ,

$$n_i = A_i \quad \text{e} \quad m_i = k_i A_i$$

onde A_i são constantes arbitrárias que serão definidas mais tarde. Quando substituídas nas soluções experimentais, essas expressões para n_i e m_i , juntamente com os valores de b_i , levarão a formas específicas de funções complementares. Se todas as raízes forem números reais distintos, podemos aplicar (18.5) e escrever

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \sum m_i b_i^t \\ \sum n_i b_i^t \end{bmatrix} = \begin{bmatrix} \sum k_i A_i b_i^t \\ \sum A_i b_i^t \end{bmatrix}$$

Contudo, no caso de raízes repetidas, devemos aplicar (18.6), e o resultado é que as funções complementares conterão termos com um t multiplicativo extra, tais como $m_1 b^t + m_2 t b^t$ (para x_c) e $n_1 b^t + n_2 t b^t$ (para y_c). Os fatores de proporcionalidade entre m_i e n_i devem ser determinados pela relação entre as variáveis x e y como estipulada nas equações do sistema dado, tal como ilustrado em (19.10) em nosso exemplo numérico. Por fim, no caso de raízes complexas, as funções complementares devem ser escritas utilizando (18.10) como protótipo.

Por fim, para obter a solução geral, podemos simplesmente formar a soma

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$$

Então, resta apenas definir as constantes arbitrárias A_i .

A extensão desse procedimento para o sistema de n equações deve ser evidente por si só. Entretanto, quando n for grande, pode não ser fácil resolver a equação característica – uma equação polinomial de n -ésimo grau – em termos quantitativos. Nesse caso, mais uma vez veremos que o teorema de Schur pode nos ajudar a chegar a certas conclusões qualitativas sobre as trajetórias temporais das variáveis no sistema. Lembramos que, nas soluções experimentais, a mesma base b é atribuída a todas essas variáveis, portanto, elas devem terminar com as mesmas expressões b_i^t nas funções complementares e compartilhar as mesmas propriedades de convergência. Assim, uma única aplicação do teorema de Schur nos habilitará a determinar a convergência ou divergência da trajetória temporal de cada uma das variáveis no sistema.

Equações diferenciais simultâneas

O método de solução que acabamos de descrever também pode ser aplicado a um sistema de equações *diferenciais* lineares de primeira ordem. Praticamente a única modificação importante que precisa ser feita é mudar as soluções experimentais para

$$x(t) = me^{rt} \quad \text{e} \quad y(t) = ne^{rt} \quad (19.11)$$

o que implica que

$$x'(t) = rme^{rt} \quad \text{e} \quad y'(t) = rne^{rt} \quad (19.12)$$

Segundo a convenção de notação que adotamos, as raízes características agora são denotadas por r em vez de b .

Suponha que temos o seguinte sistema de equações:

$$\begin{aligned} x'(t) + 2y'(t) + 2x(t) + 5y(t) &= 77 \\ y'(t) + x(t) + 4y(t) &= 61 \end{aligned} \quad (19.13)$$

Em primeiro lugar, vamos reescrevê-lo em notação matricial como

$$Ju + Mv = g \quad (19.13')$$

onde as matrizes são

$$J = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad u = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} \quad M = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \quad v = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad g = \begin{bmatrix} 77 \\ 61 \end{bmatrix}$$

Note que, em vista da aparição do termo $2y'(t)$ na primeira equação de (19.13), temos de usar a matriz J em lugar da matriz identidade I , como em (19.4''). É claro que, se J for invertível (de modo que J^{-1} exista), então podemos, em certo sentido, *normalizar* (19.13') pré-multiplicando cada um de seus termos por J^{-1} , para obter

$$J^{-1}Ju + J^{-1}Mv = J^{-1}g \quad \text{ou} \quad Iu + Kv = d \quad (K \equiv J^{-1}M; d \equiv J^{-1}g) \quad (19.13'')$$

Esse novo formato é uma duplicata exata de (19.4''), embora seja bom lembrar que os vetores u e v têm significados totalmente diferentes nos dois contextos diferentes. No desenvolvimento que se segue, vamos aderir à formulação $Ju + Mv = g$ dada em (19.13').

Para encontrar as soluções particulares, vamos tentar soluções constantes $x(t) = \bar{x}$ e $y(t) = \bar{y}$ — que implicam que $x'(t) = y'(t) = 0$. Se essas soluções forem válidas, os vetores v e u se tornarão $v = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$ e

$u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, e (19.13') será reduzida a $Mv = g$. Assim, a solução para x e y pode ser escrita como

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \bar{v} = M^{-1}g \quad (19.14)$$

a qual você deve comparar com (19.5'). Em termos numéricos, nosso problema atual dá como resultado as seguintes soluções particulares:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{5}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} 1 \\ 15 \end{bmatrix}$$

Em seguida, vamos procurar as funções complementares. Usando as soluções experimentais sugeridas em (19.11) e (19.12), os vetores u e v tornam-se

$$u = \begin{bmatrix} m \\ n \end{bmatrix} r e^{rt} \quad \text{e} \quad v = \begin{bmatrix} m \\ n \end{bmatrix} e^{rt}$$

A substituição dessas expressões na equação homogênea

$$Ju + Mv = 0$$

dá como resultado

$$J = \begin{bmatrix} m \\ n \end{bmatrix} r e^{rt} + M \begin{bmatrix} m \\ n \end{bmatrix} e^{rt} = 0$$

ou, após multiplicar tudo pelo escalar e^{-rt} e fatorar,

$$(rJ + M) \begin{bmatrix} m \\ n \end{bmatrix} = 0 \quad (19.15)$$

Você deve comparar essa expressão com (19.8'). Uma vez que nosso objetivo é achar soluções *não-triviais* de m e n (de modo que nossas soluções experimentais também serão não-triviais), é necessário que

$$|rJ + M| = 0 \quad (19.16)$$

Análoga a (19.9'), esta última equação – a equação característica do sistema de equações dado – dará como resultado as raízes r_i que precisamos. Então, podemos achar os valores correspondentes (não-triviais) de m_i e n_i .

Em nosso exemplo atual, a equação característica é

$$|rJ + M| = \begin{vmatrix} r+2 & 2r+5 \\ 1 & r+4 \end{vmatrix} = r^2 + 4r + 3 = 0 \quad (19.16')$$

com raízes $r_1 = -1$, $r_2 = -3$. Substituindo essas raízes em (19.15), obtemos

$$\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = 0 \quad (\text{para } r_1 = -1)$$

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = 0 \quad (\text{para } r_2 = -3)$$

Segue-se que $m_1 = -3n_1$ e $m_2 = -n_2$, as quais também podem ser expressas como

$$\begin{array}{ll} m_1 = 3A_1 & \text{e} \quad m_2 = A_2 \\ n_1 = -A_1 & n_2 = -A_2 \end{array}$$

Agora que r_i , m_i e n_i foram todas encontradas, as funções complementares podem ser escritas como as seguintes combinações lineares de expressões exponenciais:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \sum m_i e^{r_i t} \\ \sum n_i e^{r_i t} \end{bmatrix} \quad [\text{raízes reais distintas}]$$

E a solução geral surgirá na forma

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$$

Em nosso exemplo atual, a solução é

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 3A_1 e^{-t} + A_2 e^{-3t} + 1 \\ -A_1 e^{-t} - A_2 e^{-3t} + 15 \end{bmatrix}$$

Além disso, se tivermos as condições iniciais $x(0) = 6$ e $y(0) = 12$, podemos constatar que as constantes arbitrárias são $A_1 = 1$ e $A_2 = 2$. Elas servirão para definir a solução precedente.

Uma vez mais podemos observar que, uma vez que as expressões $e^{r_i t}$ são compartilhadas por ambas as trajetórias temporais $x(t)$ e $y(t)$, estas últimas devem ambas convergir ou ambas divergir. Como no caso atual as raízes são -1 e -3 , ambas as trajetórias convergem para os seus equilíbrios respectivos, a saber, $\bar{x} = 1$ e $\bar{y} = 15$.

Ainda que nosso exemplo consista em um sistema de duas equações apenas, o método certamente se estende ao sistema geral de n equações. Quando n for grande, novamente as soluções quantitativas podem ser difíceis, mas, uma vez encontrada a equação característica, sempre será possível uma análise qualitativa recorrendo-se ao teorema de Routh.

Comentários adicionais sobre a equação característica

Até agora o termo “equação característica” já foi encontrado em *três* contextos separados: na Seção 11.3, falamos da equação característica de uma matriz; nas Seções 16.1 e 18.1, o termo foi aplicado a uma equação diferencial linear única e a uma equação de diferenças única; agora, nesta seção, acabamos de apresentar a equação característica de um sistema de equações de diferenças ou diferenciais lineares. Há uma conexão entre os três?

Na verdade há, e a conexão é próxima. Em primeiro lugar, dada uma equação única e um sistema de equações equivalente – como exemplificado pela equação (19.1) e pelo sistema (19.1'), ou pela equação (19.3) e pelo sistema (19.3') – suas equações características devem ser idênticas. Como ilustração, considere a equação de diferenças (19.1), $y_{t+2} + a_1 y_{t+1} + a_2 y_t = c$. Já aprendemos antes a escrever sua equação característica transplantando seus coeficientes constantes diretamente para uma equação quadrática:

$$b^2 + a_1 b + a_2 = 0$$

E o sistema equivalente (19.1')? Considerando que esse sistema esteja na forma de $Iu + Kv = d$, como em (19.4''), temos a matriz $K = \begin{bmatrix} a_1 & a_2 \\ -1 & 0 \end{bmatrix}$. Assim, a equação característica é

$$|bI + K| = \begin{vmatrix} b + a_1 & a_2 \\ -1 & b \end{vmatrix} = b^2 + a_1 b + a_2 = 0 \quad [\text{por (19.9')}] \quad (19.17)$$

que é exatamente igual à obtida pela equação única, como tínhamos afirmado. Naturalmente, o mesmo tipo de resultado também vale na estrutura da equação diferencial, sendo que a única diferença é que, de acordo com nossa convenção, substituiríamos o símbolo b pelo símbolo r na última estrutura.

Também é possível ligar a equação característica de um sistema de equações de diferenças (ou diferenciais) à de uma matriz quadrada específica, que denominaremos D . Com referência à definição em (11.14), mas utilizando o símbolo b (em vez de r) para a estrutura da equação de diferenças, podemos escrever a equação característica da matriz D como se segue:

$$|D - bI| = 0 \quad (19.18)$$

Em geral, se multiplicarmos cada elemento do determinante $|D - bI|$ por -1 , o valor do determinante permanecerá inalterado se a matriz D contiver um número *par* de linhas (ou colunas) e mudará de sinal se D contiver um número *ímpar* de linhas. No presente caso, contudo, uma vez que $|D - bI|$ deve ser igualado a zero, multiplicar cada um dos elementos por -1 não terá importância, independentemente da dimensão da matriz D . Mas multiplicar cada um dos elementos do determinante $|D - bI|$ por -1 equivale a multiplicar a matriz $(D - bI)$ por -1 (veja Exemplo 6 da Seção 5.3) antes de tomar seu determinante. Assim, (19.18) pode ser reescrito como

$$|bI - D| = 0 \quad (19.18')$$

Quando essa expressão é igualada a (19.17), fica claro que, se escolhermos a matriz $D = -K$, então sua equação característica será idêntica à do sistema (19.1'). Essa matriz, $-K$, tem um significado especial: se tomarmos a versão *homogênea* do sistema, $Iu + Kv = 0$, e a expressarmos na forma de Iu

$= -Kv$, ou simplesmente $u = -Kv$, vemos que $-K$ é a matriz que pode transformar o vetor $v = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$ no vetor $u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix}$ naquela equação particular.

Mais uma vez, o mesmo raciocínio pode ser adaptado ao sistema de equações diferenciais (19.3'). Contudo, no caso de um sistema como (19.13'), $Ju + Mv = g$, onde $-$ diferentemente do sistema (19.3') – o primeiro termo é Ju e não Iu , a equação característica está na forma

$$|rJ + M| = 0 \quad [\text{cf. (19.16')}]$$

Para esse caso, se quisermos achar a expressão para a matriz D , devemos primeiramente normalizar a equação $Ju + Mv = g$ para a forma de (19.13'') e então tomar $D = -K = -J^{-1}M$.

Em suma, dados (1) uma equação de diferenças ou diferencial única e (2) um sistema de equações equivalente, do qual possamos também obter (3) uma matriz apropriada D , se tentarmos achar as equações características de todas essas três, os resultados devem ser um só e o mesmo.

EXERCÍCIO 19.2

- Verifique se o sistema de equações de diferenças (19.4) é equivalente à equação única $y_{t+2} + 6y_{t+1} + 9y_t = 4$, que foi resolvida anteriormente como o Exemplo 4 na Seção 18.1. Como se comparam as soluções obtidas pelos dois métodos diferentes?
- Mostre que a equação característica da equação de diferenças (19.2) é idêntica à do sistema equivalente (19.2').
- Resolva os dois sistemas de equações de diferenças seguintes:
 - $$\begin{aligned} (a) \quad x_{t+1} &+ x_t + 2y_t &= 24 \\ y_{t+1} + 2x_t - 2y_t &= 9 \end{aligned} \quad (\text{com } x_0 = 10 \text{ e } y_0 = 9)$$
 - $$\begin{aligned} (b) \quad x_{t+1} &- x_t - \frac{1}{3}y_t &= -1 \\ x_{t+1} + y_{t+1} &- \frac{1}{6}y_t &= 8 \frac{1}{2} \end{aligned} \quad (\text{com } x_0 = 5 \text{ e } y_0 = 4)$$
- Resolva os dois sistemas de equações diferenciais seguintes:
 - $$\begin{aligned} (a) \quad x'(t) &- x(t) - 12y(t) &= -60 \\ y'(t) + x(t) + 6y(t) &= 36 \end{aligned} \quad [\text{com } x(0) = 13 \text{ e } y(0) = 4]$$
 - $$\begin{aligned} (b) \quad x'(t) &- 2x(t) + 3y(t) &= 10 \\ y'(t) - x(t) + 2y(t) &= 9 \end{aligned} \quad [\text{com } x(0) = 8 \text{ e } y(0) = 5]$$
- Com base no sistema de equações diferenciais (19.13), encontre a matriz D cuja equação característica é idêntica à do sistema. Verifique se as equações características dos dois são realmente as mesmas.

19.3 Modelos dinâmicos de insumo-produto

Na primeira vez que tratamos da análise insumo-produto, a pergunta era: quanto deve ser produzido em cada indústria de modo que os requisitos de insumos de todas as indústrias, bem como a demanda final (sistema aberto), sejam exatamente satisfeitos? O contexto era estático e o problema era resolver um sistema de equações simultâneas para os níveis de *equilíbrio* da produção de todas as indústrias. Quando certas considerações econômicas adicionais são incorporadas ao modelo, o sistema insumo-produção pode assumir um caráter dinâmico e então resultará em um sistema de equações de diferenças – ou diferenciais – do tipo discutido na Seção 19.2.

Aqui serão abordadas três dessas considerações dinamizadoras. Para manter a simplicidade da explicação, contudo, ilustraremos com sistemas abertos de duas indústrias, apenas. Ainda assim, uma vez que vamos empregar notação matricial, a generalização para o caso de n indústrias não deve ser difícil, pois ela pode ser conseguida simplesmente mudando as dimensões das matrizes envolvidas como devido. Para o propósito de tal generalização, será aconselhável denotar as variáveis por $x_{1,t}$ e $x_{2,t}$ em vez de x_t e y_t , de modo que possamos estender a notação a $x_{n,t}$ quando necessário. Você recordará que, no contexto do insumo-produção, x_i representa a produção (medida em dólares) da i -ésima indústria; agora, o novo índice t adicionará a essa produção uma dimensão de tempo. O símbolo de coeficiente de insumo-produção a_{ij} ainda significará o valor em dólares da i -ésima mercadoria requerida na produção do valor de um dólar da j -ésima mercadoria e d_i novamente indicará a demanda final pela i -ésima mercadoria.

Defasagem de tempo na produção

Em um sistema aberto estático de duas indústrias, o produto da indústria I deve ser estabelecido no nível da demanda da seguinte maneira:

$$x_1 = a_{11}x_1 + a_{12}x_2 + d_1$$

Agora suponha uma defasagem de um período na produção, de modo que a quantia demandada no período t determine não a produção corrente, mas a produção do período $(t+1)$. Para retratar essa nova situação, devemos modificar a equação precedente para a forma

$$x_{1,t+1} = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} \quad (19.19)$$

De modo semelhante, podemos escrever, para a indústria II:

$$x_{2,t+1} = a_{21}x_{1,t} + a_{22}x_{2,t} + d_{2,t} \quad (19.19')$$

Desse modo, agora temos um sistema de equações de diferenças simultâneas, o que constitui uma versão dinâmica do modelo de insumo-produção.

Em notação matricial, o sistema consiste na equação

$$x_{t+1} - Ax_t = d_t \quad (19.20)$$

$$\text{onde} \quad x_{t+1} = \begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} \quad x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d_t = \begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}$$

É claro que (19.20) está na forma de (19.4''), com somente duas exceções. A primeira é que, diferentemente do vetor u , o vetor x_{t+1} não tem como seu “coeficiente” uma matriz identidade I . Contudo, como já explicado antes, do ponto de vista analítico isso não faz realmente nenhuma diferença. O segundo ponto, mais substantivo, é que o vetor d_t com um índice de tempo, implica que o vetor demanda final está sendo considerado como uma função de tempo. Se essa função não for constante, será preciso fazer uma modificação no método para encontrar as soluções par-

ticulares, embora as funções complementares permaneçam inalteradas. O exemplo seguinte ilustrará o procedimento modificado.

EXEMPLO 1 Dado o vetor demanda final exponencial

$$d_t = \begin{bmatrix} \delta^t \\ \delta^t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t \quad (\delta = \text{um escalar positivo})$$

encontre as soluções particulares do modelo dinâmico de insumo-produto (19.20). De acordo com o método de coeficientes indeterminados apresentado na Seção 18.4, devemos tentar soluções da forma $x_{1,t} = \beta_1 \delta^t$ e $x_{2,t} = \beta_2 \delta^t$, onde β_1 e β_2 são coeficientes indeterminados. Isto é, devemos tentar

$$x_t = \begin{bmatrix} \beta_1 \delta^t \\ \beta_2 \delta^t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t \quad (19.21)$$

que implica[†]

$$x_{t+1} = \begin{bmatrix} \beta_1 \delta^{t+1} \\ \beta_2 \delta^{t+1} \end{bmatrix} = \begin{bmatrix} \beta_1 \delta \\ \beta_2 \delta \end{bmatrix} \delta^t = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t$$

Se as soluções experimentais indicadas forem válidas, então o sistema (19.20) se tornará

$$\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t - \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t$$

ou, cancelando o multiplicador escalar comum $\delta^t \neq 0$,

$$\begin{bmatrix} \delta - a_{11} & -a_{12} \\ -a_{21} & \delta - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (19.22)$$

Supondo que a matriz de coeficientes na extrema esquerda seja invertível, podemos constatar, de imediato (pela regra de Cramer), que β_1 e β_2 são

$$\beta_1 = \frac{\delta - a_{22} + a_{12}}{\Delta} \quad \text{e} \quad \beta_2 = \frac{\delta - a_{11} + a_{21}}{\Delta} \quad (19.22')$$

onde $\Delta \equiv (\delta - a_{11})(\delta - a_{22}) - a_{12}a_{21}$. Visto que β_1 e β_2 agora são expressas inteiramente nos valores conhecidos dos parâmetros, basta inseri-las na solução experimental (19.21) para obter as expressões definidas para as soluções particulares.

Uma versão mais geral do tipo de vetor demanda final discutido aqui é dada no Exercício 19.3–1.

O procedimento para encontrar as funções complementares de (19.20) não é diferente do apresentado na Seção 19.2. Uma vez que a versão homogênea do sistema de equações é $x_{t+1} - Ax_t = 0$, a equação característica deve ser

$$|bI - A| = \begin{vmatrix} b - a_{11} & -a_{12} \\ -a_{21} & b - a_{22} \end{vmatrix} = 0 \quad [\text{cf. (19.9')}]$$

[†] Você notará que o vetor $\begin{bmatrix} \beta_1 \delta \\ \beta_2 \delta \end{bmatrix}$ pode ser reescrito de diversas formas equivalentes:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta \quad \text{ou} \quad \delta \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{ou} \quad \delta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Aqui escolhemos a terceira alternativa porque, em uma etapa subsequente, vamos adicionar $\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}$ a uma outra matriz 2×2 . As duas primeiras formas alternativas acarretarão problemas de conformidade de dimensão.

Por essa expressão, podemos encontrar as raízes características b_1 e b_2 e então passar para as etapas restantes do processo de solução.

Excesso de demanda e ajuste da produção

A formulação do modelo em (19.20) também pode surgir de uma premissa econômica diferente. Considere a situação na qual o excesso de demanda para cada produto sempre tende a induzir um incremento de produção igual ao excesso de demanda. Visto que o excesso de demanda para o primeiro produto no período t equivale a

$$\underbrace{a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t}}_{\text{demandado}} - \underbrace{x_{1,t}}_{\text{fornecido}}$$

o ajuste da produção (incremento) $\Delta x_{1,t}$ deve ser estabelecido exatamente igual àquele nível:

$$\Delta x_{1,t} (\equiv x_{1,t+1} - x_{1,t}) = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} - x_{1,t}$$

Contudo, se somarmos $x_{1,t}$ a ambos os lados dessa equação, o resultado se tornará idêntico a (19.19). De modo semelhante, nossa premissa de ajuste de produção nos dará uma equação igual a (19.19') para a segunda indústria. Resumindo, o mesmo modelo matemático pode resultar de premissas econômicas totalmente diferentes.

Até aqui, o sistema insumo-produção tem sido considerado somente na estrutura do tempo discreto. Com a finalidade de comparação, agora vamos utilizar o molde do tempo contínuo para o processo de ajuste da produção.

Em essência, isso exigiria a utilização do símbolo $x_i(t)$ em lugar de $x_{i,t}$ e da derivada $x'_i(t)$ em lugar da diferença $\Delta x_{i,t}$. Com essas mudanças, nossa premissa de ajuste da produção se manifestará só no seguinte par de equações diferenciais:

$$\begin{aligned} x'_1(t) &= a_{11}x_1(t) + a_{12}x_2(t) + d_1(t) - x_1(t) \\ x'_2(t) &= a_{21}x_1(t) + a_{22}x_2(t) + d_2(t) - x_2(t) \end{aligned}$$

A qualquer instante de tempo $t = t_0$, o símbolo $x_i(t_0)$ nos informa a taxa de fluxo de produção por unidade de tempo (digamos, por mês) que prevalece no instante em questão, e $d_i(t_0)$ indica a demanda final por mês prevalecente naquele instante. Logo, a soma no lado direito de cada equação indica a taxa de excesso de demanda por mês, medida no instante $t = t_0$. Por outro lado, a derivada $x'_i(t_0)$ no lado esquerdo representa a taxa de ajuste da produção mensal exigida pelo excesso de demanda em $t = t_0$. Esse ajuste erradicará o excesso de demanda (e efetuará o equilíbrio) dentro de um mês, mas somente se ambos, excesso de demanda e ajuste da produção, permanecerem inalterados às taxas correntes. Na verdade, o excesso de demanda variará ao longo do tempo, assim como o ajuste da produção induzido, resultando em uma caçada de gato e rato. A solução do sistema, que consiste nas trajetórias temporais da produção x_i , narra a história dessa caçada. Se a solução for convergente, no devido tempo o gato (ajuste da produção) acabará capturando o rato (excesso de demanda), assintoticamente (quando $t \rightarrow \infty$).

Após rearranjo adequado, esse sistema de equações diferenciais pode ser escrito no formato de (19.13'), como se segue:

$$I x' + (I - A)x = d \quad (19.23)$$

$$\text{onde } x' = \begin{bmatrix} x'_1(t) \\ x'_2(t) \end{bmatrix} \quad x = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad d = \begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix}$$

(a “linha” denota derivada, e não transposição). As funções complementares podem ser encontradas pelo método discutido anteriormente. Em particular, as raízes características devem ser obtidas pela equação

$$|rI + (I - A)| = \begin{vmatrix} r+1-a_{11} & -a_{12} \\ -a_{21} & r+1-a_{12} \end{vmatrix} = 0 \quad [\text{cf. (19.16)}]$$

Quanto às soluções particulares, se o vetor demanda final contiver, como seus elementos, funções não-constantes de tempo $d_1(t)$ e $d_2(t)$, será necessária uma modificação no método de solução. Vamos ilustrar com um exemplo simples.

EXEMPLO 2 Dado o vetor demanda final

$$d = \begin{bmatrix} \lambda_1 e^{\rho t} \\ \lambda_2 e^{\rho t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} e^{\rho t}$$

onde λ_i e ρ são constantes, encontre as soluções particulares do modelo dinâmico (19.23). Utilizando o método de coeficientes indeterminados, podemos tentar soluções da forma $x_i(t) = \beta_i e^{\rho t}$, que implicam, é claro, que $x'_i(t) = \rho \beta_i e^{\rho t}$. Em notação matricial, essas expressões podem ser escritas como

$$x = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \quad (19.24)$$

$$\text{e } x' = \rho \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \quad [\text{cf. nota de esclarecimento no Exemplo 1}]$$

Após substituir em (19.23) e cancelar o multiplicador escalar comum (diferente de zero) $e^{\rho t}$, obtemos

$$\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1-a_{11} & -a_{12} \\ -a_{21} & 1-a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

ou

$$\begin{bmatrix} \rho+1-a_{11} & -a_{12} \\ -a_{21} & \rho+1-a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \quad (19.25)$$

Se a matriz mais à esquerda for invertível, podemos aplicar a regra de Cramer e determinar que os valores dos coeficientes β_i são

$$\begin{aligned} \beta_1 &= \frac{\lambda_1(\rho+1-a_{22}) + \lambda_2 a_{12}}{\Delta} \\ \beta_2 &= \frac{\lambda_2(\rho+1-a_{11}) + \lambda_1 a_{21}}{\Delta} \end{aligned} \quad (19.25')$$

onde $\Delta \equiv (\rho+1-a_{11})(\rho+1-a_{22}) - a_{12}a_{21}$. Agora que os *coeficientes indeterminados* foram determinados, podemos introduzir esses valores na solução experimental (19.24) para obter as soluções particulares desejadas.

Formação de capital

Uma outra consideração econômica que pode dar origem a um sistema dinâmico insumo-produto é a formação de capital, incluindo a acumulação de estoque.

Na discussão estática, consideramos somente o nível de produção de cada produto necessário para satisfazer a demanda corrente. As necessidades de acumulação de estoque ou de formação de capital foram ignoradas ou incorporadas ao vetor demanda final. Para expor a formação de capital, agora vamos considerar – juntamente com uma matriz de coeficientes de insumos $A = [a_{ij}]$ – uma matriz de coeficientes de capital

$$C = [c_{ij}] = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

onde c_{ij} denota o valor em dólar da i -ésima mercadoria que a j -ésima indústria necessita como novo capital (seja equipamento ou estoque, dependendo da natureza da i -ésima mercadoria), resultante de um incremento de produção de \$1 na j -ésima indústria. Por exemplo, se um aumento de \$1 na produção da indústria de refrigerantes (j -ésima) induzi-la a acrescentar o valor de \$2 em equipamento de engarrafamento (i -ésima mercadoria), então $c_{ij} = 2$. Assim, tal coeficiente de capital revela uma certa razão marginal capital/produção, sendo a razão limitada a um só tipo de capital apenas (a i -ésima mercadoria). Assim como os coeficientes de insumo a_{ij} , supomos que os coeficientes de capital são fixos. A idéia é que a economia produza cada mercadoria em quantidade tal que satisfaça não somente a demanda de necessidade de insumo mais a demanda final, mas também a demanda de necessidade de capital para tal.

Se o tempo for *contínuo*, o incremento de produção é indicado pelas derivadas $x'_i(t)$; assim, a produção de cada indústria deve ser estabelecida em

$$\begin{aligned} x_1(t) &= a_{11}x_1(t) + a_{12}x_2(t) + c_{11}x'_1(t) + c_{12}x'_2(t) + d_1(t) \\ x_2(t) &= \underbrace{a_{21}x_1(t) + a_{22}x_2(t)}_{\text{necessidade de insumo}} + \underbrace{c_{21}x'_1(t) + c_{22}x'_2(t)}_{\text{necessidade de capital}} + \underbrace{d_2(t)}_{\text{demanda final}} \end{aligned}$$

Em notação matricial, isso pode ser expresso pela equação

$$I x = A x + C x' + d$$

ou

$$C x' + (A - I)x = -d \quad (19.26)$$

Se o tempo for *discreto*, a necessidade de capital no período t será baseada no incremento da produção $x_{i,t} - x_{i,t-1}$ ($\equiv \Delta x_{i,t-1}$); assim, os níveis de produção devem ser estabelecidos em

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}}_{\text{necessidade de insumo}} + \underbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} - x_{1,t-1} \\ x_{2,t} - x_{2,t-1} \end{bmatrix}}_{\text{necessidade de capital}} + \underbrace{\begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}}_{\text{demanda final}}$$

ou

$$I x_t = A x_t + C(x_t - x_{t-1}) + d_t$$

Adiantando os índices de tempo por um período e reunindo termos, podemos escrever a equação na forma

$$(I - A - C)x_{t+1} + Cx_t = d_{t+1} \quad (19.27)$$

Mais uma vez, o sistema de equações diferenciais (19.26) e o sistema de equação de diferenças (19.27) podem ser resolvidos, é claro, pelo método da Seção 19.2. E também fica subentendido que essas duas equações matriciais são ambas extensíveis para o caso de n indústrias por uma simples redefinição adequada das matrizes e uma mudança correspondente em suas dimensões.

Acabamos de discutir como um modelo dinâmico de insumo-produção pode se originar de considerações como defasagens de tempo e mecanismos de ajuste. Quando considerações semelhantes são aplicadas a modelos de mercado de equilíbrio geral, esses modelos tendem a se tornar dinâmicos praticamente do mesmo modo. Mas, visto que o espírito da formulação de tais modelos é análogo ao dos modelos de insumo-produto, prescindiremos de uma discussão formal e aconselhamos que você consulte os casos ilustrativos nos Exercícios 19.3-6 e 19.3-7.

EXERCÍCIO 19.3

- No Exemplo 1, se o vetor demanda final for mudado para $d_t = \begin{bmatrix} \lambda_1 \delta^t \\ \lambda_2 \delta^t \end{bmatrix}$, quais serão as soluções particulares? Após encontrar suas respostas, mostre que as respostas no Exemplo 1 são um mero caso especial dessas, com $\lambda_1 = \lambda_2 = 1$.
- Mostre que (19.22) pode ser escrita, com maior concisão, como $(\delta I - A)\beta = u$
 - Dos cinco símbolos utilizados, quais são escalares? Quais são vetores? Quais são matrizes?
 - Escreva a solução para β em forma de matriz, supondo que $(\delta I - A)$ é invertível.
- Mostre que (19.25) pode ser escrita, com maior concisão, como $(\rho I + I - A)\beta = \lambda$
 - Quais dos cinco símbolos representam, respectivamente, escalares, vetores e matrizes?
 - Escreva a solução para β em forma de matriz, supondo que $(\rho I + I - A)$ é invertível.
- Dados $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ e $d_t = \begin{bmatrix} \left(\frac{12}{10}\right)^t \\ \left(\frac{12}{10}\right)^t \end{bmatrix}$ para o modelo de insumo-produção de tempo discreto com defasagem de tempo descrito em (19.20), encontre (a) as soluções particulares; (b) as funções complementares; e (c) as trajetórias temporais definidas, supondo produções iniciais $x_{1,0} = \frac{187}{39}$ e $x_{2,0} = \frac{72}{13}$. (Use frações, e não decimais, em todos os cálculos).
- Dados $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ e $d = \begin{bmatrix} e^{t/10} \\ 2e^{t/10} \end{bmatrix}$ para o modelo de insumo-produção de tempo contínuo com ajuste de produção descrito em (19.23), encontre (a) as soluções particulares; (b) as funções complementares; e (c) as trajetórias temporais definidas, supondo as condições iniciais $x_1(0) = \frac{53}{6}$ e $x_2(0) = \frac{25}{6}$. (Use frações, e não decimais, em todos os cálculos.)
- Em um mercado de n mercadorias, todas as Q_{di} e Q_{si} (com $i = 1, 2, \dots, n$) podem ser consideradas como funções dos n preços $P_1, \dots, P_n$, assim como o excesso de demanda para cada mercadoria $E_i \equiv Q_{di} - Q_{si}$. Supondo linearidade, podemos escrever

$$E_1 = a_{10} + a_{11}P_1 + a_{12}P_2 + \dots + a_{1n}P_n$$

$$E_2 = a_{20} + a_{21}P_1 + a_{22}P_2 + \dots + a_{2n}P_n$$

$$\dots$$

$$E_n = a_{n0} + a_{n1}P_1 + a_{n2}P_2 + \dots + a_{nn}P_n$$
 ou, em notação matricial,

$$E = a + AP$$
 - O que estes últimos quatro símbolos representam – escalares, vetores ou matrizes? Quais são suas respectivas dimensões?
 - Considere que todos os preços são funções de tempo e suponha que $dP_i/dt = \alpha_i E_i$ ($i = 1, 2, \dots, n$). Qual é a interpretação econômica deste último conjunto de equações?
 - Escreva as equações diferenciais mostrando que cada dP_i/dt é uma função linear dos n preços.
 - Mostre que, se denotarmos por P' o vetor coluna $n \times 1$ das derivadas dP_i/dt , e se denotarmos por α uma matriz diagonal $n \times n$, com $\alpha_1, \alpha_2, \dots, \alpha_n$ (nessa ordem) na diagonal principal e zeros em todos os outros lugares, podemos escrever o sistema de equações diferenciais precedente em notação matricial como $P' - \alpha AP = \alpha a$.
- Para o mercado de n mercadorias do Problema 6, a versão de tempo discreto consistiria em um conjunto de equações de diferenças $\Delta P_{i,t} = \alpha_i E_{i,t}$ ($i = 1, 2, \dots, n$), onde $E_{i,t} = a_{i0} + a_{i1}P_{1,t} + a_{i2}P_{2,t} + \dots + a_{in}P_{n,t}$.
 - Escreva o sistema de equações de excesso de demanda e mostre que ele pode ser expresso em notação matricial como $E_t = a + AP_t$.
 - Mostre que as equações de ajuste de preço podem ser escritas como $P_{t+1} - P_t = \alpha E_t$ onde α é a matriz diagonal $n \times n$ definida no Problema 6.
 - Mostre que o sistema de equação de diferenças do presente modelo de tempo discreto pode ser expresso na forma $P_{t+1} - (I + \alpha A)P_t = \alpha a$.

19.4 Mais uma vez, o modelo de inflação-desemprego

Agora que já ilustramos o tipo multissetor de sistemas dinâmicos com modelos de insumo-produção, vamos dar um exemplo de modelo econômico de equações dinâmicas simultâneas no cenário de um só setor. Para essa finalidade, o modelo inflação-desemprego, que já encontramos duas vezes sob aspectos diferentes, pode ser chamado ao serviço mais uma vez.

Equações diferenciais simultâneas

Na Seção 16.5, o modelo de inflação-desemprego foi apresentado na estrutura de tempo contínuo por meio das três equações seguintes:

$$p = \alpha - T - \beta U + g\pi \quad (\alpha, \beta > 0; 0 < g \leq 1) \quad (16.33)$$

$$\frac{d\pi}{dt} = j(p - \pi) \quad (0 < j \leq 1) \quad (16.34)$$

$$\frac{dU}{dt} = -k(\mu - p) \quad (k > 0) \quad (16.35)$$

exceto que aqui, adotamos a letra grega μ no lugar de m em (16.35) para evitar confusão com a utilização anterior do símbolo m na discussão metodológica da Seção 19.2. Quando tratamos desse modelo na Seção 16.5, uma vez que ainda não estávamos aptos para lidar com equações dinâmicas simultâneas, enfrentamos o problema condensando o modelo em uma equação única de uma só variável, o que exigia um processo bastante trabalhoso de substituições e eliminações. Agora, em vista da coexistência de dois padrões de variação dados no modelo para π e U , trataremos o modelo como um de duas equações diferenciais simultâneas.

Quando (16.33) é substituída nas outras duas equações e as derivadas $d\pi/dt \equiv \pi'(t)$ e $dU/dt \equiv U'(t)$ são escritas de modo mais simples como π' e U' , o modelo toma a forma

$$\begin{aligned} \pi' + j(1 - g)\pi + j\beta U &= j(\alpha - T) \\ U' - kg\pi + k\beta U &= k(\alpha - T - \mu) \end{aligned} \quad (19.28)$$

ou, em notação matricial,

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_J \underbrace{\begin{bmatrix} \pi' \\ U' \end{bmatrix}}_M + \underbrace{\begin{bmatrix} j(1 - g) & j\beta \\ -kg & k\beta \end{bmatrix}}_M \underbrace{\begin{bmatrix} \pi \\ U \end{bmatrix}}_M = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix} \quad (19.28')$$

Por esse sistema, as trajetórias temporais de π e U podem ser encontradas simultaneamente. Então, se desejado, podemos obter a trajetória p utilizando (16.33).

Trajетórias de solução

Para encontrar as soluções particulares, podemos simplesmente estabelecer $\pi' = U' = 0$ (fazer π e U constantes ao longo do tempo) em (19.28') e resolver para π e U . Em nossa discussão anterior, em (19.14), essas soluções foram obtidas por meio da inversão de matriz, mas, certamente, a regra de Cramer também pode ser usada. Seja como for, podemos constatar que

$$\bar{\pi} = \mu \quad \text{e} \quad \bar{U} = \frac{1}{\beta} [\alpha - T - (1 - g)\mu] \quad (19.29)$$

O resultado $\bar{\pi} = \mu$ (a taxa de inflação esperada no equilíbrio é igual à taxa de expansão monetária) coincide com o alcançado na Seção 16.5. Quanto à taxa de desemprego U , nenhuma tentati-

va foi feita para encontrar seu nível de equilíbrio naquela seção. Entretanto, se a tivéssemos feito (com base na equação diferencial em U dada no Exercício 16.5–2), a resposta não seria diferente da solução $\bar{U}$ em (19.29).

Voltando às funções complementares, que são baseadas nas soluções experimentais me^{rt} e ne^{rt} , podemos determinar m , n e r pela equação matricial homogênea

$$(rJ + M) \begin{bmatrix} m \\ n \end{bmatrix} = 0 \quad [\text{de (19.15)}]$$

que, no presente contexto, toma a forma de

$$\begin{bmatrix} r + j(1 - g) & j\beta \\ -kg & r + k\beta \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (19.30)$$

Para evitar soluções triviais para m e n a partir desse sistema homogêneo, devemos anular o determinante da matriz de coeficientes, ou seja, é preciso que

$$|rJ + M| = r^2 + [k\beta + j(1 - g)]r + k\beta j = 0 \quad (19.31)$$

Essa equação quadrática, uma versão específica da equação característica $r^2 + a_1 r + a_2 = 0$, tem coeficientes

$$a_1 = k\beta + j(1 - g) \quad \text{e} \quad a_2 = k\beta j$$

E esses coeficientes, como seria de esperar, são exatamente os valores a_1 e a_2 em (16.37'') – uma versão de uma equação única do presente modelo na variável π . O resultado é que a análise anterior dos três casos de raízes características deve-se aplicar aqui com igual validade. Entre outras conclusões, podemos recordar que, independentemente de as raízes serem reais ou complexas, a parte real de cada raiz no presente modelo é sempre negativa. Assim, as trajetórias de solução são sempre convergentes.

EXEMPLO 1 Encontre as trajetórias temporais de π e U , dados os valores de parâmetros

$$\alpha - T = \frac{1}{6} \quad \beta = 3 \quad g = 1 \quad j = \frac{3}{4} \quad \text{e} \quad k = \frac{1}{2}$$

Uma vez que esses valores de parâmetros são os mesmos que os do Exemplo 1 na Seção 16.5, os resultados da presente análise podem ser verificados de imediato em comparação com os resultados da seção citada.

Em primeiro lugar, é fácil determinar que as soluções particulares são

$$\bar{\pi} = \mu \quad \text{e} \quad \bar{U} = \frac{1}{3} \left(\frac{1}{6} \right) = \frac{1}{18} \quad [\text{por (19.29)}] \quad (19.32)$$

Como a equação característica é

$$r^2 + \frac{3}{2}r + \frac{9}{8} = 0 \quad [\text{por (19.31)}]$$

temos que as duas raízes são complexas:

$$r_1, r_2 = \frac{1}{2} \left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}} \right) = -\frac{3}{4} \pm \frac{3}{4}i \quad \left(\text{com } h = -\frac{3}{4} \text{ e } v = \frac{3}{4} \right) \quad (19.33)$$

A substituição das duas raízes (juntamente com os valores de parâmetros) dá como resultado, respectivamente, as equações matriciais

$$\begin{bmatrix} -\frac{3}{4}(1-i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1+i) \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \left[de \ r_1 = -\frac{3}{4} + \frac{3}{4}i \right] \quad (19.34)$$

$$\begin{bmatrix} -\frac{3}{4}(1+i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1-i) \end{bmatrix} \begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \left[de \ r_2 = -\frac{3}{4} - \frac{3}{4}i \right] \quad (19.34')$$

Visto que r_1 e r_2 são projetadas – via (19.31) – para tornar singular a matriz de coeficientes, cada uma das duas equações matriciais precedentes contém, na verdade, somente uma equação independente, que pode determinar apenas uma relação de proporcionalidade entre as constantes arbitrárias m_i e n_i . Especificamente, temos

$$\frac{1}{3}(1-i)m_1 = n_1 \quad e \quad \frac{1}{3}(1+i)m_2 = n_2$$

De acordo com isso, as funções complementares podem ser expressas como

$$\begin{aligned} \begin{bmatrix} \pi_c \\ U_c \end{bmatrix} &= \begin{bmatrix} m_1 e^{r_1 t} + m_2 e^{r_2 t} \\ n_1 e^{r_1 t} + n_2 e^{r_2 t} \end{bmatrix} \\ &= e^{ht} \begin{bmatrix} m_1 e^{vit} + m_2 e^{-vit} \\ n_1 e^{vit} + n_2 e^{-vit} \end{bmatrix} \quad [\text{por (16.11)}] \\ &= e^{ht} \begin{bmatrix} (m_1 + m_2) \cos vt + (m_1 - m_2) i \sin vt \\ (n_1 + n_2) \cos vt + (n_1 - n_2) i \sin vt \end{bmatrix} \quad [\text{por (16.24)}] \end{aligned}$$

Se, por questão de simplicidade de notação, definirmos novas constantes arbitrárias

$$A_5 \equiv m_1 + m_2 \quad e \quad A_6 \equiv (m_1 - m_2)i$$

então segue-se que[†]

$$n_1 + n_2 = \frac{1}{3}(A_5 - A_6) \quad (n_1 - n_2)i = \frac{1}{3}(A_5 + A_6)$$

Portanto, usando essas expressões e incorporando os valores h e v de (19.33) nas funções complementares, acabamos com

[†] Isso pode ser visto pelo seguinte:

$$\begin{aligned} n_1 + n_2 &= \frac{1}{3}(1-i)m_1 + \frac{1}{3}(1+i)m_2 = \frac{1}{3}[(m_1 + m_2) - (m_1 - m_2)i] \\ &= \frac{1}{3}(A_5 - A_6) \\ (n_1 - n_2)i &= \left[\frac{1}{3}(1-i)m_1 - \frac{1}{3}(1+i)m_2 \right] i = \frac{1}{3}[(m_1 - m_2) - (m_1 + m_2)i]i \\ &= \frac{1}{3}(A_6 + A_5) \quad [i^2 \equiv -1] \end{aligned}$$

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = e^{-3t/4} \begin{bmatrix} A_5 \cos \frac{3}{4}t + A_6 \sin \frac{3}{4}t \\ \frac{1}{3}(A_5 - A_6) \cos \frac{3}{4}t + \frac{1}{3}(A_5 + A_6) \sin \frac{3}{4}t \end{bmatrix} \quad (19.35)$$

Por fim, combinando as soluções particulares em (19.32) com as funções complementares acima, podemos obter as trajetórias de solução de π e U . Como seria de esperar, essas trajetórias são exatamente as mesmas que as trajetórias em (16.43) e (16.45) na Seção 16.5.

Equações de diferenças simultâneas

O tratamento de equações simultâneas para o modelo inflação-desemprego em tempo discreto é de natureza semelhante à discussão precedente de tempo contínuo. Assim, tocaremos apenas nos pontos relevantes.

O modelo em questão, como dado na Seção 18.3, consiste em três equações, duas das quais descrevem os padrões de variação de π e U , respectivamente:

$$p_t = \alpha - T - \beta U_t + g\pi_t \quad (18.18)$$

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \quad (18.19)$$

$$U_{t+1} - U_t = -k(\mu - p_{t+1}) \quad (18.20)$$

Eliminando p e reunindo termos, podemos reescrever o modelo como o sistema de equações de diferenças

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -kg & 1 + \beta k \end{bmatrix}}_J \underbrace{\begin{bmatrix} \pi_{t+1} \\ U_{t+1} \end{bmatrix}}_K + \underbrace{\begin{bmatrix} -(1-j+jg) & j\beta \\ 0 & -1 \end{bmatrix}}_K \underbrace{\begin{bmatrix} \pi_t \\ U_t \end{bmatrix}}_K = \underbrace{\begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix}}_J \quad (19.36)$$

Trajetórias de solução

Se existirem equilíbrios constantes, as soluções particulares de (19.36) podem ser expressas como $\bar{\pi} = \pi_t = \pi_{t+1}$ e $\bar{U} = U_t = U_{t+1}$. Substituindo $\bar{\pi}$ e $\bar{U}$ em (19.36) e resolvendo o sistema (por inversão de matriz ou pela regra de Cramer), obtemos

$$\bar{\pi} = \mu \quad \text{e} \quad \bar{U} = \frac{1}{\beta} [\alpha - T - (1-g)\mu] \quad (19.37)$$

O valor de $\bar{U}$ é o mesmo que o valor encontrado na Seção 18.3. Embora não tenhamos obtido $\bar{\pi}$ na última seção, as informações no Exercício 18.3-2 indicam que $\bar{\pi} = \mu$, o que está de acordo com (19.37). De fato, como você pode notar, os resultados em (19.37) também são idênticos aos valores de equilíbrio intertemporal obtidos na estrutura de tempo contínuo em (19.29).

A busca pelas funções complementares, dessa vez com base nas soluções experimentais mb^t e nb^t , envolve a equação matricial homogênea

$$(bJ + K) \begin{bmatrix} m \\ n \end{bmatrix} = 0$$

ou, em vista de (19.36),

$$\begin{bmatrix} b - (1-j+jg) & j\beta \\ -bkg & b(1+\beta k) - 1 \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (19.38)$$

Para evitar soluções triviais desse sistema homogêneo, é preciso que

$$|bf + K| = (1 + \beta k)b^2 - [1 + gj + (1 - j)(1 + \beta k)]b + (1 - j + jg) = 0 \quad (19.39)$$

A versão normalizada dessa equação quadrática é a equação característica $b^2 + a_1b + a_2 = 0$, com os mesmos coeficientes a_1 e a_2 de (18.24) e (18.33) na Seção 18.3. Por consequência, a análise dos três casos de raízes características realizadas naquela seção devem se aplicar igualmente aqui.

Para cada raiz, b_p , (19.38) nos fornece uma relação de proporcionalidade específica entre as constantes arbitrárias m_i e n_i , e essas relações nos permitem ligar as constantes arbitrárias existentes na função complementar para U às existentes na função complementar para π . Então, combinando as funções complementares e as soluções particulares, podemos obter as trajetórias temporais de π e U .

EXERCÍCIO 19.4

1. Verifique (19.29) usando a regra de Cramer.
2. Verifique se surge a mesma relação de proporcionalidade entre m_1 e n_1 quer usemos a primeira ou a segunda equação no sistema (19.34).
3. Encontre as trajetórias temporais (soluções gerais) de π e U , dados:

$$p = \frac{1}{6} - 2U + \frac{1}{3}\pi$$

$$\pi' = \frac{1}{4}(p - \pi)$$

$$U' = -\frac{1}{2}(\mu - p)$$

4. Encontre as trajetórias temporais (soluções gerais) de π e U , dados:

$$(a) \quad p_t = \frac{1}{2} - 3U_t + \frac{1}{2}\pi_t$$

$$(b) \quad p_t = \frac{1}{4} - 4U_t + \pi_t$$

$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$

$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$

$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

19.5 Diagramas de fase de duas variáveis

As seções precedentes trataram de soluções *quantitativas* de sistemas dinâmicos *lineares*. Nesta seção, discutiremos a análise *gráfico-qualitativa* (diagrama de fase) de um sistema de equações diferenciais *não-linear*. Mais especificamente, nossa atenção estará voltada para o sistema de equações diferenciais de primeira ordem com duas variáveis, na forma geral de

$$\begin{aligned} x'(t) &= f(x, y) \\ y'(t) &= g(x, y) \end{aligned}$$

Note que as derivadas de temporais $x'(t)$ e $y'(t)$ dependem somente de x e y e que a variável t não entra nas funções f e g como um argumento separado. Essa característica, que transforma o sistema em um *sistema autônomo*, é um pré-requisito para a aplicação da técnica do diagrama de fase.[†]

[†] No diagrama de fase de uma variável apresentado anteriormente na Seção 15.6, a equação $dy/dt = f(y)$ também é restrita a ser *autônoma*, sendo que é proibido ter a variável t como um argumento explícito na função f .

O diagrama de fase de duas variáveis, assim como a versão de uma variável na Seção 15.6, é limitado no sentido de que pode responder apenas a questões qualitativas – as concernentes à localização e à estabilidade dinâmica do(s) equilíbrio(s) intertemporal(is). Mas, novamente, assim como a versão de uma variável, ele tem as vantagens compensadoras de poder tratar sistemas não-lineares do mesmo modo confortável que trata sistemas lineares e enfrentar problemas enunciados em termos de funções gerais tão facilmente quanto os enunciados em termos específicos.

O espaço de fase

Quando construímos o diagrama de fase de uma variável (Figura 15.3) para a equação diferencial (autônoma) $dy/dt = f(y)$, simplesmente fizemos o gráfico de dy/dt em função de y nos dois eixos em um espaço de fase bidimensional. Contudo, agora que o número de variáveis é *dobrado*, o que podemos fazer para atender à aparente necessidade de mais eixos? Felizmente, a resposta é que o espaço bidimensional é tudo que precisamos.

Para ver por que isso é viável, observe que a tarefa mais crucial da construção do diagrama de fase é determinar a direção do movimento da(s) variável(is) ao longo do tempo. É essa informação, incorporada às setas da Figura 15.3, que nos permite obter as inferências qualitativas finais. Para desenhar essas setas, precisamos somente de duas coisas: (1) uma linha de demarcação – vamos denominá-la linha “ $dy/dt = 0$ ” – que fornece o lugar para qualquer equilíbrio (ou equilíbrios) prospectivo e, o que é mais importante, divide o espaço de fase em duas regiões, uma caracterizada por $dy/dt > 0$ e a outra por $dy/dt < 0$; e (2) uma reta real sobre a qual os aumentos e diminuições de y que são implicados por quaisquer valores diferentes de zero de dy/dt podem ser indicados. Na Figura 15.3, a linha de demarcação citada no item 1 é encontrada no eixo horizontal. Mas, na realidade, esse eixo também serve como a linha real citada no item 2. Isso significa que, para dy/dt , na verdade podemos abandonar o eixo vertical sem nenhum prejuízo, contanto que tomemos o cuidado de distinguir entre a região $dy/dt > 0$ e a região $dy/dt < 0$ – digamos, marcando a primeira com um sinal de mais, e a última com um sinal de menos. O fato de podermos prescindir de um eixo é que torna viável a colocação de um diagrama de fase de duas variáveis no espaço bidimensional. Agora precisamos de *duas* retas reais, em vez de uma. Mas isso é automaticamente resolvido pelos eixos x e y , padrões em um diagrama bidimensional. Agora também precisamos de *duas* linhas de demarcação (ou curvas), uma para $dx/dt = 0$ e a outra para $dy/dt = 0$. Mas elas podem ser representadas graficamente em um espaço de fase de duas dimensões. E, tão logo desenhadas, não seria difícil decidir quais lados dessas duas linhas ou curvas devem ser marcados com sinais de mais e menos, respectivamente.

As curvas de demarcação

Dado o seguinte sistema de equações diferenciais autônomo

$$\begin{aligned} x' &= f(x, y) \\ y' &= g(x, y) \end{aligned} \quad (19.40)$$

onde x' e y' são a forma abreviada para as derivadas temporais $x'(t)$ e $y'(t)$, respectivamente, as duas curvas de demarcação – que serão denotadas por $x' = 0$ e $y' = 0$ – representam os gráficos das duas equações

$$f(x, y) = 0 \quad [\text{curva } x' = 0] \quad (19.41)$$

$$g(x, y) = 0 \quad [\text{curva } y' = 0] \quad (19.42)$$

Se a forma específica da função f for conhecida, (19.41) pode ser resolvida para y em termos de x e a solução pode ser representada graficamente no plano xy como a curva $x' = 0$. Contudo, ainda que não seja, mesmo assim podemos recorrer à regra da função implícita e determinar que a inclinação da curva $x' = 0$ é

$$\left. \frac{dy}{dx} \right|_{x'=0} = - \frac{\partial f / \partial x}{\partial f / \partial y} = - \frac{f_x}{f_y} \quad (f_y \neq 0) \quad (19.43)$$

Contanto que os sinais das derivadas parciais f_x e f_y ($\neq 0$) sejam conhecidos, é possível obter uma indicação qualitativa da inclinação da curva $x' = 0$ por (19.43). Pelo mesmo critério, a inclinação da curva $y' = 0$ pode ser inferida da derivada

$$\left. \frac{dy}{dx} \right|_{y'=0} = - \frac{g_x}{g_y} \quad (g_y \neq 0) \quad (19.44)$$

Para uma ilustração mais concreta, vamos supor que

$$f_x < 0 \quad f_y > 0 \quad g_x > 0 \quad \text{e} \quad g_y < 0 \quad (19.45)$$

Então, ambas as curvas $x' = 0$ e $y' = 0$ terão inclinações positivas. Se supusermos ainda mais que

$$-\frac{f_x}{f_y} > -\frac{g_x}{g_y} \quad [\text{a curva } x' = 0 \text{ é mais inclinada que a curva } y' = 0]$$

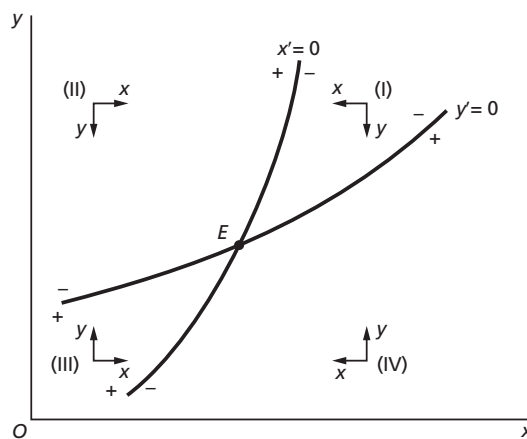
então podemos encontrar uma situação tal qual a mostrada na Figura 19.1. Note que as linhas de demarcação agora podem ser curvas. Note também que, agora, elas já não precisam coincidir com os eixos.

As duas curvas de demarcação, que se interceptam no ponto E , dividem o espaço de fase em quatro regiões distintas, rotuladas I a IV. O ponto E , onde x e y são ambos constantes ($x' = y' = 0$), representa o equilíbrio intertemporal do sistema. Contudo, em qualquer outro ponto, ou x ou y (ou ambos) estariam variando ao longo do tempo, em direções determinadas pelos sinais das derivadas de tempo x' e y' naquele ponto. Na presente instância, acontece que temos $x' > 0$ ($x' < 0$) à esquerda (direita) da curva $x' = 0$; daí os sinais de mais (menos) à esquerda (direita) daquela curva. Esses sinais se baseiam no fato de que

$$\frac{\partial x'}{\partial x} = f_x < 0 \quad [\text{por (19.40) e (19.45)}] \quad (19.46)$$

o que implica que, à medida que passamos continuamente de oeste para leste no espaço de fase (à medida que x aumenta), x' sofre uma redução constante, de modo que o sinal de x' deve passar por três estágios, na ordem +, 0, -. De modo análogo, a derivada

FIGURA 19.1



$$\frac{\partial y'}{\partial y} = g_y < 0 \quad [\text{por (19.40) e (19.45)}] \quad (19.47)$$

implica que, à medida que passamos continuamente do sul para o norte (à medida que y aumenta), y' decresce constantemente, de modo que o sinal de y' deve passar por três estágios, na ordem $+$, 0 , $-$. Assim, somos levados a acrescentar os sinais de mais abaixo da curva $y' = 0$ na Figura 19.1, e os sinais de menos acima dessa curva.

Com base nesses sinais de mais e menos, agora podemos desenhar um conjunto de setas direcionais para indicar o movimento intertemporal de x e y . Para qualquer ponto na região I, x' e y' são ambas negativas. Por conseguinte, ambos, x e y , devem decrescer com o tempo, produzindo um movimento na *direção oeste* para x e um movimento na *direção sul* para y . Como indicam as duas setas na região I, dado um ponto inicial localizado na região I, o movimento intertemporal deve ser na direção geral sudoeste. O exato oposto é verdade na região III, onde x' e y' são ambas positivas, de modo que ambas as variáveis x e y devem crescer ao longo do tempo. Por outro lado, x' e y' têm sinais diferentes na região II. Com x' positiva e y' negativa, x deve se mover na *direção leste* e y na *direção sul*. E a região IV apresenta uma tendência exatamente oposta à da região II.

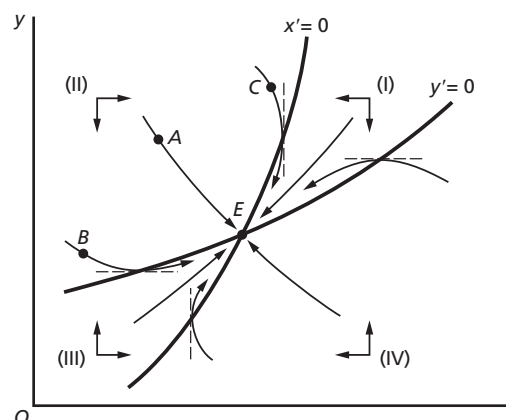
Linhas de fluxo

Para entender melhor as implicações das setas direcionais, podemos esboçar uma série de *linhas de fluxo* no diagrama de fase. Também conhecidas como *caminhos de fase* ou simplesmente *caminhos* ou *trajetórias de fase*, essas linhas de fluxo servem para visualizar o movimento dinâmico do sistema a partir de qualquer ponto inicial concebível. Alguns desses pontos estão ilustrados na Figura 19.2, que reproduz as curvas $x' = 0$ e $y' = 0$ na Figura 19.1. Uma vez que cada ponto no espaço de fase deve estar localizado sobre uma linha de fluxo ou outra, deve existir um número infinito de linhas de fluxo, todas elas conforme os requisitos direcionais impostos pelas setas xy em cada uma das regiões. Contudo, para retratar o caráter qualitativo geral do diagrama de fase, normalmente algumas poucas linhas de fluxo representativas serão suficientes.

Podemos notar diversas características das linhas de fluxo na Figura 19.2. Em primeiro lugar, acontece que todas elas levam ao ponto E , o que faz desse ponto um equilíbrio intertemporal estável (aqui, globalmente estável). Mais tarde encontraremos outros tipos de configurações de linhas de fluxo. Em segundo lugar, enquanto algumas linhas de fluxo nunca se aventuram para além de uma única região (tal como a que passa pelo ponto A), outras podem cruzar de uma região para outra (tais como as que passam por B e C). Em terceiro, onde uma linha de fluxo cruzar, ela deve ter ou uma inclinação infinita (cruzando a curva $x' = 0$) ou uma inclinação zero (cruzando a curva $y' = 0$). Isso se deve ao fato de que ao longo da curva $x' = 0$ ($y' = 0$), $x(y)$ é estacionário ao longo do tempo, portanto a linha de fluxo não deve ter nenhum movimento horizontal (vertical) ao cruzar aquela curva. Para assegurar que esses requisitos de inclinação sejam consistentemente cumpridos, tão logo as curvas de demarcação tenham sido posicionadas, seria aconselhável adicionar algumas barras *verticais* curtas através da curva $x' = 0$ e algumas barras *horizontais* através da curva $y' = 0$, para servir como diretrizes no desenho das linhas de fluxo.[†] Em quarto e último lugar, embora as linhas de fluxo indiquem explicitamente as direções do movimento de x e y ao longo do tempo, elas não dão nenhuma informação específica em relação à velocidade e à aceleração, porque o diagrama de fase não permite a presença de um eixo t (tempo). É por essa razão, é claro, que as linhas de fluxo ganharam o nome alternativo *trajetórias de fase* e não *trajetórias temporais*. A única observação que podemos fazer sobre a velocidade é de natureza qualitativa: à medida que nos movemos ao longo de uma linha de fluxo cada vez mais próximos da curva $x' = 0$ ($y' = 0$), a velocidade de aproximação na direção horizontal (vertical) deve diminuir progressivamente. Isso se deve à redução constante do valor absoluto da derivada $x' \equiv dx/dt$ ($y' \equiv dy/dt$), que ocorre à medida que nos aproximamos da linha de demarcação na qual $x'(y')$ assume o valor zero.

[†] Como recurso mnemônico, note que as barras auxiliares que atravessam a curva $x' = 0$ devem ser *perpendiculares* ao eixo x . De modo semelhante, as barras auxiliares que atravessam a curva $y' = 0$ devem ser *perpendiculares* ao eixo y .

FIGURA 19.2



Tipos de equilíbrio

Dependendo das configurações das linhas de fluxo que circundam um determinado equilíbrio intemporal, esse equilíbrio pode cair em uma de quatro categorias: (1) nós; (2) pontos de sela; (3) focos; e (4) vórtices.

Um *nó* é um equilíbrio tal que todas as linhas de fluxo associadas a ele ou fluem de modo não-cíclico em sua direção (*nó estável*) ou fluem de modo não-cíclico afastando-se dele (*nó instável*). Já encontramos um nó estável na Figura 19.2. Um nó instável é mostrado na Figura 19.3a. Note que, nessa ilustração particular, acontece que as linhas de fluxo nunca cruzam de região para região. Além disso, acontece que as curvas $x' = 0$ e $y' = 0$ são retas e, de fato, elas mesmas servem como linhas de fluxo.

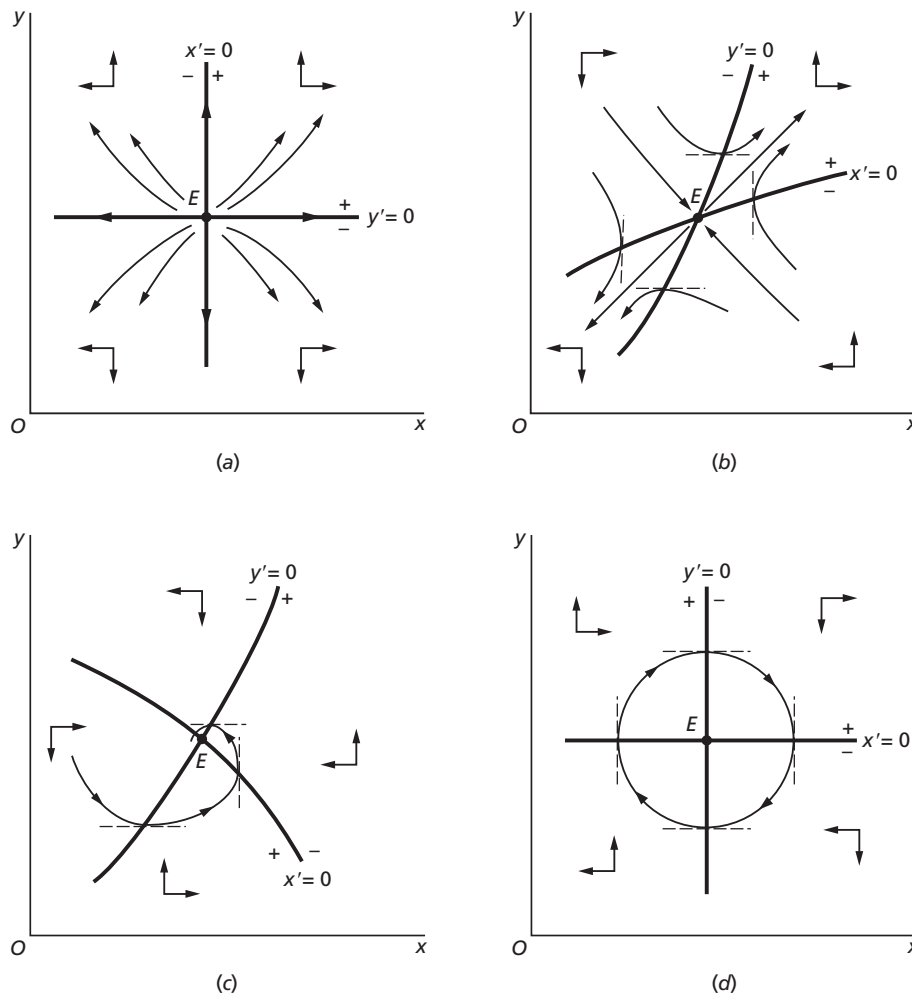
Um *ponto de sela* é um equilíbrio que tem dupla personalidade – é estável em algumas direções, mas instável em outras. Mais exatamente, com referência à ilustração na Figura 19.3b, um ponto de sela tem exatamente um par de linhas de fluxo – denominadas os *ramos estáveis* do ponto de sela – que fluem de modo direto e consistente em direção ao equilíbrio, e exatamente um par de linhas de fluxo – os *ramos instáveis* – que fluem de modo direto e consistente afastando-se dele. Todas as outras trajetórias dirigem-se inicialmente ao ponto de sela, mas, cedo ou tarde, se afastam dele. É claro que foi essa dupla personalidade que inspirou o nome “ponto de sela”. Uma vez que observa-se estabilidade somente nos ramos estáveis que, aliás, não podem ser alcançados, um ponto de sela é classificado genericamente como um equilíbrio *instável*.

O terceiro tipo de equilíbrio, *foco*, é caracterizado por trajetórias giratórias, todas as quais fluem em ciclos em direção a ele (*foco estável*) ou afastando-se dele, também em ciclos (*foco instável*). A Figura 19.3c ilustra um foco estável com apenas uma linha de fluxo desenhada explicitamente para evitar confusão. O que causa a ocorrência do movimento giratório? A resposta está no modo como as curvas $x' = 0$ e $y' = 0$ estão posicionadas. Na Figura 19.3c, as duas curvas de demarcação têm inclinações tais que cada uma por vez bloqueia a linha de fluxo que corre na direção prescrita por um conjunto específico de setas xy . O resultado é que a linha de fluxo é freqüentemente obrigada a cruzar de uma região para a outra, traçando uma espiral. Obter um foco estável (como é o caso aqui) ou um foco instável depende do posicionamento relativo das duas curvas de demarcação. Mas, em qualquer caso, a inclinação da linha de fluxo nos pontos de cruzamento ainda deve ser ou infinita (cruzando $x' = 0$) ou zero (cruzando $y' = 0$).

Por fim, podemos ter um *vórtice* (ou *centro*). Mais uma vez, este é um equilíbrio com linhas de fluxo giratórias, mas agora elas formam uma família de circuitos completos (círculos ou ovais concêntricos) que orbitam ao redor do equilíbrio em movimento perpétuo. Um exemplo disso é dado na Figura 19.3d, na qual, novamente, mostramos apenas uma linha de fluxo. Considerando que esse tipo de equilíbrio é inatingível a partir de qualquer posição afastada do ponto E, um vórtice é automaticamente classificado como um equilíbrio *instável*.

Todas as ilustrações na Figura 19.3 exibem um equilíbrio único. Contudo, quando existe não-linearidade suficiente, as duas curvas de demarcação podem se interceptar mais de uma vez e, por conseguinte, produzir múltiplos equilíbrios. Quando isso acontece, pode existir, no mes-

FIGURA 19.3



mo diagrama de fase, uma combinação dos tipos de equilíbrio intertemporal citados anteriormente. Embora nesse caso haja mais do que quatro regiões para tratar, o princípio subjacente da análise de diagrama de fase sempre será basicamente o mesmo.

Inflação e regra monetária à moda de Obst

Como ilustração econômica do diagrama de fase de duas variáveis, apresentaremos um modelo devido ao professor Obst,[†] cuja intenção é mostrar a ineficácia do tipo convencional de regra contracíclica de política monetária (e, portanto, a necessidade de uma nova regra) quando está em vigor um “mecanismo de ajuste da inflação”. Esse modelo contrasta com o que discutimos anteriormente sobre inflação, no sentido de que, em vez de estudar as implicações de uma taxa de expansão monetária *dada*, ele vai adiante e examina a eficácia de duas *regras* monetárias diferentes, cada uma prescrevendo um conjunto diferente de ações monetárias, a serem executadas em face de várias condições inflacionárias.

Uma premissa crucial do modelo é o mecanismo de ajuste da inflação

$$\frac{dp}{dt} = h \left(\frac{M_s - M_d}{M_s} \right) = h \left(1 - \frac{M_d}{M_s} \right) \quad (h > 0) \quad (19.48)$$

[†] Obst, Norman P. “Stabilization Policy with an Inflation Adjustment Mechanism”. *Quarterly Journal of Economics*, p. 355–359, maio 1978. Em seu artigo, Obst não apresenta nenhum diagrama de fase, mas eles podem ser construídos sem nenhuma dificuldade a partir do modelo.

que mostra que o efeito de um excesso de oferta de moeda ($M_s > M_d$) é elevar a taxa de inflação p , em vez do nível de preço P . Assim, a compensação do mercado monetário não implicaria estabilidade, mas apenas um taxa estável de inflação. Para facilitar a análise, a segunda igualdade em (19.48) serve para deslocar o foco do excesso de oferta de moeda para a razão demanda-oferta de moeda, M_d/M_s , que denotaremos por μ . Tendo como premissa que M_d é diretamente proporcional ao produto nacional nominal PQ , podemos escrever

$$\mu \equiv \frac{M_d}{M_s} = \frac{aPQ}{M_s} \quad (a > 0)$$

Então, as taxas de crescimento das diversas variáveis estão relacionadas por

$$\frac{d\mu/dt}{\mu} = \frac{da/dt}{a} + \frac{dP/dt}{P} + \frac{dQ/dt}{Q} - \frac{dM_s/dt}{M_s} \quad (19.49)$$

[por (10.24) e (10.25)]

$$\equiv p + q - m \quad [a = \text{uma constante}]$$

onde as letras minúsculas p , q e m denotam, respectivamente, a taxa de inflação, a taxa (exógena) de crescimento do produto nacional real e a taxa de expansão monetária.

As equações (19.48) e (19.49), um conjunto de duas equações diferenciais, podem determinar, em conjunto, as trajetórias temporais de p e μ se, por enquanto, considerarmos m como exógena. Usando os símbolos p' e μ' para representar as derivadas temporais $p'(t)$ e $\mu'(t)$, podemos expressar esse sistema com mais concisão como

$$\begin{aligned} p' &= h(1 - \mu) \\ \mu' &= (p + q - m)\mu \end{aligned} \quad (19.50)$$

Dado que h é positivo, podemos ter $p' = 0$ se, e somente se, $1 - \mu = 0$. De modo semelhante, uma vez que μ é sempre positivo, $\mu' = 0$ se, e somente se, $p + q - m = 0$. Assim, as curvas de demarcação $p' = 0$ e $\mu' = 0$ estão associadas às equações

$$\mu = 1 \quad [\text{curva } p' = 0] \quad (19.51)$$

$$p = m - q \quad [\text{curva } \mu' = 0] \quad (19.52)$$

Como mostra a Figura 19.4a, a representação gráfica dessas expressões é uma reta horizontal e uma reta vertical e dá como resultado um equilíbrio único em E . O valor de equilíbrio $\bar{\mu} = 1$ significa que, em equilíbrio, M_d e M_s são iguais, compensando o mercado monetário. O fato de se demonstrar que a taxa de inflação de equilíbrio é positiva reflete uma premissa implícita de que $m > q$.

Uma vez que a curva $p' = 0$ corresponde à curva $x' = 0$ em nossa discussão anterior, ela deve ter barras auxiliares verticais. E a outra curva deve ter barras auxiliares horizontais. Por (19.50), constatamos que

$$\frac{\partial p'}{\partial \mu} = -h < 0 \quad \text{e} \quad \frac{\partial \mu'}{\partial p} = \mu > 0 \quad (19.53)$$

com a implicação de que um movimento na direção norte através da curva $p' = 0$ passa pela sequência de sinais $(+, 0, -)$ para p' , e um movimento na direção leste através da curva $\mu' = 0$ passa pela sequência de sinais $(-, 0, +)$ para μ' . Assim, obtemos os quatro conjuntos de setas direcionais, tal como desenhado, que geram linhas de fluxo (somente uma delas é mostrada) que orbitam ao

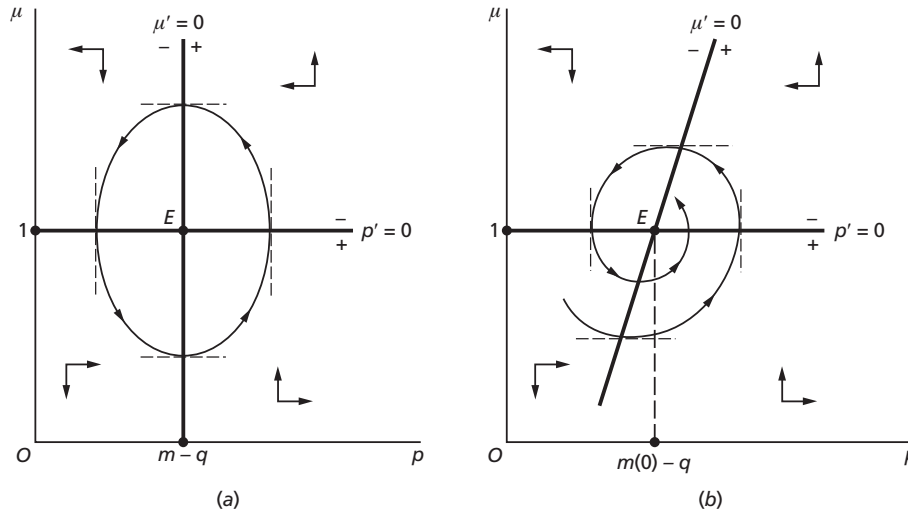


FIGURA 19.4

CHIANG / WAINWRIGHT

redor do ponto E em sentido anti-horário. É claro que isso faz de E um vórtice. A menos que aconteça de a economia estar inicialmente em E , é impossível obter o equilíbrio. Ao contrário, haverá uma flutuação incessante.

Contudo, a conclusão precedente é consequência de uma taxa de expansão monetária *exógena*. E se agora transformarmos m em endógena adotando uma regra monetária antiinflacionária? A regra monetária “convencional” exigiria que a taxa de expansão monetária engrenasse negativamente com a taxa de inflação:

$$m = m(p) \quad m'(p) < 0 \quad [\text{regra monetária convencional}] \quad (19.54)$$

Tal regra modificaria a segunda equação em (19.50) para

$$\mu' = [p + q - m(p)]\mu \quad (19.55)$$

e alteraria (19.52) para

$$p = m(p) - q \quad [\text{curva } \mu' = 0 \text{ sob a regra monetária convencional}] \quad (19.56)$$

Dado que $m(p)$ é monotônica, existe somente um valor de p – digamos, p_1 – que pode satisfazer essa equação. Por conseguinte, a curva $\mu' = 0$ ainda deve surgir como uma reta vertical, embora com uma interseção horizontal diferente, $p_1 = m(p_1) - q$. Além do mais, por (19.55), constatamos que

$$\frac{\partial \mu'}{\partial p} = [1 - m'(p)]\mu > 0 \quad [\text{por (19.54)}]$$

que, em termos *qualitativos*, não é diferente da derivada em (19.53). Deduz-se que as setas direcionais também devem permanecer como estão na Figura 19.4a. Em suma, acabaríamos com um vórtice, como antes.

A regra monetária alternativa proposta por Obst é engrenar m à taxa de variação (e não ao nível) da taxa de inflação:

$$m = m(p') \quad m'(p') < 0 \quad [\text{regra monetária alternativa}] \quad (19.57)$$

Sob essa regra, (19.55) e (19.56) se tornarão, respectivamente,

$$\mu' = [p + q - m(p')]\mu \quad (19.58)$$

$$p = m(p') - q \quad [\text{curva } \mu' = 0 \text{ sob a regra monetária alternativa}] \quad (19.59)$$

Dessa vez, a inclinação da curva $\mu' = 0$ se tornaria ascendente. Pois, diferenciando (19.59) em relação a μ utilizando a regra da cadeia, temos

$$\frac{dp}{d\mu} = m'(p') \frac{dp'}{d\mu} = m'(p') (-h) > 0 \quad [\text{por (19.50)}]$$

portanto, pela regra da função inversa, $d\mu/dp$ – a inclinação da curva $\mu' = 0$ – também é positiva. Essa nova situação é ilustrada na Figura 19.4b, na qual, por simplicidade, a curva $\mu' = 0$ é desenhada como uma linha reta, com uma inclinação designada arbitrariamente.[†] Apesar da mudança da inclinação, a derivada parcial

$$\frac{\partial \mu'}{\partial p} = \mu > 0 \quad [\text{por (19.58)}]$$

não mudou em relação a (19.53), portanto, as setas μ devem conservar sua orientação original na Figura 19.4a. As linhas de fluxo (das quais somente uma é mostrada) agora girarão para dentro na direção do equilíbrio em $\bar{\mu} = 1$ e $\bar{p} = m(0) - q$, onde $m(0)$ denota $m(p')$ calculado em $p' = 0$. Assim, a regra monetária alternativa é considerada capaz de converter um vórtice em um foco estável, tornando possível, desse modo, a eliminação assintótica da flutuação perpétua na taxa de inflação. De fato, com uma curva $\mu' = 0$ suficientemente reta é até possível transformar o vórtice em um nó estável.

EXERCÍCIO 19.5

- Mostre que o diagrama de fase de duas variáveis também pode ser usado se o modelo consistir em uma única equação diferencial de segunda ordem, $y''(t) = f(y', y)$, em vez de duas equações de primeira ordem.
- Os sinais de mais e de menos anexados aos dois lados das curvas $x' = 0$ e $y' = 0$ na Figura 19.1 são baseados nas derivadas parciais $\partial x'/\partial x$ e $\partial y'/\partial y$, respectivamente. As mesmas conclusões podem ser obtidas das derivadas $\partial x'/\partial y$ e $\partial y'/\partial x$?
- Usando a Figura 19.2, verifique que, se uma linha de fluxo não tiver uma inclinação infinita (zero) ao cruzar a curva $x' = 0$ ($y' = 0$), ela necessariamente violará as restrições direcionais impostas pelas setas xy .
- Como casos especiais do sistema de equações diferenciais (19.40), suponha que

(a) $f_x = 0$	$f_y > 0$	$g_x > 0$	e	$g_y = 0$
(b) $f_x = 0$	$f_y < 0$	$g_x < 0$	e	$g_y = 0$

 Construa, para cada caso, um diagrama de fase apropriado, desenhe as linhas de fluxo e determine a natureza do equilíbrio.
- (a) Mostre que é possível produzir *ou* um nó estável *ou* um foco estável a partir do sistema de equações diferenciais (19.40), se

$f_x < 0$	$f_y > 0$	$g_x < 0$	e	$g_y < 0$
-----------	-----------	-----------	---	-----------

 (b) Qual característica (ou características) especial na construção de seu diagrama é responsável pela diferença nos resultados (nó *versus* foco)?
- Com referência ao modelo de Obst, verifique que, se fizermos com que a curva $\mu' = 0$ com inclinação positiva na Figura 19.4b seja suficientemente reta, as linhas de fluxo, embora ainda caracterizadas por cruzamentos, convergirão para o equilíbrio ao modo de um nó, e não de um foco.

[†] A inclinação é inversamente proporcional ao valor absoluto de $m'(p')$. Quanto maior for a sensibilidade com que se faz a taxa de expansão monetária responder à variação da taxa de inflação p' , mais plana será a curva $\mu' = 0$ na Figura 19.4b.

19.6 Linearização de um sistema de equações diferenciais não-linear

Uma outra técnica qualitativa de analisar um sistema de equações diferenciais *não-linear* é extrair inferências pela *aproximação linear* àquele sistema, a ser obtida por meio da expansão de Taylor do sistema dado ao redor de seu equilíbrio.[†] Aprendemos, na Seção 9.5, que uma aproximação linear (ou até mesmo um polinômio de ordem mais alta) de uma função arbitrária $\phi(x)$ pode nos dar o exato valor de $\phi(x)$ no ponto de expansão, mas acarretará erros de aproximação cada vez maiores à medida que nos afastamos cada vez mais do ponto de expansão. O mesmo é válido para a aproximação linear de um sistema não-linear. No ponto de expansão – aqui, o ponto de equilíbrio E –, a aproximação linear tem exatamente o mesmo equilíbrio do sistema não-linear original. Além disso, em uma vizinhança suficientemente pequena de E , a aproximação linear deve ter a mesma configuração geral de linhas de fluxo que o sistema original. Por conseguinte, contanto que estejamos dispostos a limitar nossas inferências de estabilidade à vizinhança imediata do equilíbrio, a aproximação linear poderia servir como uma fonte adequada de informações. Tal análise, denominada *análise de estabilidade local*, pode ser usada quer por si só, quer como um suplemento da análise do diagrama de fase. Vamos tratar apenas do caso de duas variáveis.

Expansão de Taylor e linearização

Dada uma função arbitrária (sucessivamente diferenciável) de uma variável $\phi(x)$, a expansão de Taylor ao redor de um ponto x_0 dá a série

$$\phi(x) = \phi(x_0) + \phi'(x_0)(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \dots$$

$$+ \frac{\phi^{(n)}(x_0)}{n!}(x - x_0)^n + R_n$$

onde um polinômio envolvendo várias potências de $(x - x_0)$ aparece à direita. Uma estrutura semelhante caracteriza a expansão de Taylor de uma função de duas variáveis $f(x, y)$ ao redor de qualquer ponto (x_0, y_0) . Entretanto, com duas variáveis em cena, o polinômio resultante abrangeria várias potências de $(y - y_0)$, bem como de $(x - x_0)$ –, na verdade, o mesmo aconteceria com os produtos dessas duas expressões:

$$\begin{aligned} f(x, y) = & f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ & + \frac{1}{2!} [f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) \\ & + f_{yy}(x_0, y_0)(y - y_0)^2] + \dots + R_n \end{aligned} \quad (19.60)$$

Note que os coeficientes das expressões $(x - x_0)$ e $(y - y_0)$ agora são derivadas *parciais* de f , todas calculadas no ponto de expansão (x_0, y_0) .

Pela série de Taylor de uma função, a aproximação linear – ou, abreviadamente, *linearização* – é obtida pelo simples descarte de todos os termos de ordem maior que um. Assim, para o caso de apenas uma variável, a linearização é a seguinte função linear de x :

$$\phi(x_0) + \phi'(x_0)(x - x_0)$$

De modo semelhante, a linearização de (19.60) é a seguinte função linear de x e y :

$$f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

[†] No caso de múltiplos equilíbrios, cada equilíbrio requer uma aproximação linear separada.

Além disso, substituindo o símbolo de função g por f nesse resultado, também podemos obter a correspondente linearização de $g(x, y)$. Deduz-se que, dado o sistema não-linear

$$\begin{aligned}x' &= f(x, y) \\y' &= g(x, y)\end{aligned}\tag{19.61}$$

sua linearização ao redor do ponto de expansão (x_0, y_0) pode ser escrita como

$$\begin{aligned}x' &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\y' &= g(x_0, y_0) + g_x(x_0, y_0)(x - x_0) + g_y(x_0, y_0)(y - y_0)\end{aligned}\tag{19.62}$$

Se as formas específicas das funções f e g forem conhecidas, então é possível atribuir valores específicos a todas as $f(x_0, y_0)$, $f_x(x_0, y_0)$, $f_y(x_0, y_0)$ e suas contrapartes para a função g e o sistema linear (19.62) pode ser resolvido em termos quantitativos. Contudo, mesmo que as funções f e g sejam dadas em formas *gerais*, a análise quantitativa ainda é possível, bastando que os sinais de f_x , f_y , g_x e g_y sejam verificáveis.

A linearização homogênea

Para finalidades de análise de estabilidade local, a linearização (19.62) pode ser colocada em uma forma mais simples. Em primeiro lugar, visto que nosso ponto de expansão deve ser o ponto de equilíbrio $(\bar{x}, \bar{y})$, devemos substituir (x_0, y_0) por $(\bar{x}, \bar{y})$. De modo mais substantivo, uma vez que, por definição, temos, no ponto de equilíbrio, $x' = y' = 0$, deduz-se que

$$f(\bar{x}, \bar{y}) = g(\bar{x}, \bar{y}) = 0 \quad [\text{por (19.61)}]$$

portanto, o primeiro termo do lado direito de cada equação em (19.62) pode ser descartado. Fazendo essas mudanças e então multiplicando os termos remanescentes no lado direito de (19.62) e rearranjando, obtemos uma outra versão da linearização:

$$\begin{aligned}x' - f_x(\bar{x}, \bar{y})x - f_y(\bar{x}, \bar{y})y &= -f_x(\bar{x}, \bar{y})\bar{x} - f_y(\bar{x}, \bar{y})\bar{y} \\y' - g_x(\bar{x}, \bar{y})x - g_y(\bar{x}, \bar{y})y &= -g_x(\bar{x}, \bar{y})\bar{x} - g_y(\bar{x}, \bar{y})\bar{y}\end{aligned}\tag{19.63}$$

Note que, em (19.63), cada termo à direita do sinal de igual representa uma constante. Tomamos o cuidado de separar esses termos constantes de modo que possamos descartar todos eles para chegar às equações homogêneas da linearização. O resultado, que pode ser escrito em notação matricial como

$$\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\tag{19.64}$$

constitui a *linearização homogênea* de (19.61). Considerando que a análise qualitativa depende exclusivamente de conhecer as raízes características, que, por sua vez, dependem somente das equações homogêneas de um sistema, (19.64) é tudo que precisamos para a análise de estabilidade local desejada.

Avançando um pouco mais, podemos observar que a única propriedade diferente da linearização homogênea está na matriz de derivadas parciais – a matriz jacobiana do sistema não-linear (19.61) – calculada no ponto de equilíbrio $(\bar{x}, \bar{y})$. Por conseguinte, em última instância, a estabilidade ou instabilidade local do equilíbrio depende exclusivamente da composição do citado jacobiano. Por questão de conveniência de notação na discussão a seguir, denotaremos o jacobiano calculado no ponto de equilíbrio por J_E e seus elementos por a , b , c e d :

$$J_E \equiv \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}\tag{19.65}$$

Vamos supor que as duas equações diferenciais são funcionalmente independentes. Então, sempre teremos $|J_E| \neq 0$. (Para alguns casos em que $|J_E| = 0$, veja Exercício 19.6–4.)

Análise da estabilidade local

De acordo com (19.16) e utilizando (19.65), a equação característica da linearização homogênea deve ser

$$\begin{vmatrix} r - a & -b \\ -c & r - d \end{vmatrix} = r^2 - (a + d)r + (ad - bc) = 0$$

É claro que as raízes características dependem criticamente das expressões $(a + d)$ e $(ad - bc)$. A última é simplesmente o determinante do jacobiano em (19.65):

$$ad - bc = |J_E|$$

E a primeira, representando a *soma* dos elementos da diagonal principal daquele jacobiano, é denominada o *traço* de J_E , simbolizado por $\text{tr } J_E$:

$$a + d = \text{tr } J_E$$

De acordo com isso, as raízes características podem ser expressas como

$$r_1, r_2 = \frac{\text{tr } J_E \pm \sqrt{(\text{tr } J_E)^2 - 4|J_E|}}{2}$$

As grandezas relativas de $(\text{tr } J_E)^2$ e $4|J_E|$ determinarão se as duas raízes são reais ou complexas, isto é, se as trajetórias temporais de x e y são constantes ou flutuantes. Para verificar a estabilidade dinâmica do equilíbrio, por outro lado, precisamos averiguar os sinais algébricos das duas raízes. Para essa finalidade, as duas relações seguintes provarão ser de grande utilidade:

$$r_1 + r_2 = \text{tr } J_E \quad [cf. (16.5) \text{ e } (16.6)] \quad (19.66)$$

$$r_1 r_2 = |J_E| \quad (19.67)$$

Caso 1 $(\text{tr } J_E)^2 > 4|J_E|$ Nesse caso, as raízes são reais e distintas e nenhuma flutuação é possível. Por conseguinte, o ponto de equilíbrio pode ser ou um nó ou um ponto de sela, mas nunca um foco ou vórtice. Em vista de $r_1 \neq r_2$, existem três possibilidades distintas de combinação de sinal: ambas as raízes negativas, ambas as raízes positivas e duas raízes com sinais opostos.[†] Levando em conta as informações em (19.66) e (19.67), essas três possibilidades são caracterizadas por:

$$\begin{aligned} (i) \quad r_1 < 0, r_2 < 0 & \Rightarrow |J_E| > 0; \text{tr } J_E < 0 \\ (ii) \quad r_1 > 0, r_2 > 0 & \Rightarrow |J_E| > 0; \text{tr } J_E > 0 \\ (iii) \quad r_1 > 0, r_2 < 0 & \Rightarrow |J_E| < 0; \text{tr } J_E \geq 0 \end{aligned}$$

Sob a Possibilidade *i*, com duas raízes negativas, ambas as funções complementares x_c e y_c tendem a zero quando t se torna infinito. Assim, o ponto de equilíbrio é um nó estável. O oposto é verda-

[†] Uma vez que excluímos $|J_E| = 0$, nenhuma raiz pode assumir um valor igual a zero.

deiro sob a Possibilidade *ii*, que descreve um nó instável. Por outro lado, com duas raízes de sinais opostos, a Possibilidade *iii* resulta em um ponto de sela.

Para que este último caso fique mais claro, lembre-se que as funções complementares das duas variáveis sob o Caso 1 tomam a forma geral

$$\begin{aligned}x_c &= A_1 e^{r_1 t} + A_2 e^{r_2 t} \\ y_c &= k_1 A_1 e^{r_1 t} + k_2 A_2 e^{r_2 t}\end{aligned}$$

onde as constantes arbitrárias A_1 e A_2 devem ser determinadas a partir das condições iniciais. Se as condições iniciais forem tais que $A_1 = 0$, a raiz positiva r_1 sairá de cena, deixando para a raiz negativa r_2 a incumbência de fazer com que o equilíbrio seja estável. Essas condições iniciais são pertinentes aos pontos localizados nos ramos estáveis do ponto de sela. Por outro lado, se as condições iniciais forem tais que $A_2 = 0$, a raiz negativa r_2 desaparecerá de cena, deixando para a raiz positiva r_1 a incumbência de fazer com que o equilíbrio seja instável. Tais condições iniciais referem-se aos pontos que estão sobre os ramos instáveis. Considerando que todas as outras condições iniciais também envolvem $A_1 \neq 0$, todas elas devem dar origem a funções complementares divergentes. Assim, a Possibilidade *iii* dá como resultado um ponto de sela.

Caso 2 ($\text{tr } J_E)^2 = 4|J_E|$) Como neste caso as raízes são repetidas, apenas duas possibilidades de combinação de sinais podem surgir:

$$(iv) \quad r_1 < 0, r_2 < 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E < 0$$

$$(v) \quad r_1 > 0, r_2 > 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E > 0$$

Essas duas possibilidades são meras duplicatas das Possibilidades *i* e *ii*. Assim, elas indicam um nó estável e um nó instável, respectivamente.

Caso 3 ($\text{tr } J_E)^2 < 4|J_E|$) Desta vez, com raízes complexas $h \pm vi$, está presente uma flutuação crítica e temos de encontrar ou um foco ou um vórtice. Tomando como base (19.66) e (19.67), temos no presente caso,

$$\begin{aligned}\text{tr } J_E &= r_1 + r_2 = (h + vi) + (h - vi) = 2h \\ |J_E| &= r_1 r_2 = (h + vi)(h - vi) = h^2 + v^2\end{aligned}$$

Assim, $\text{tr } J_E$ tem de adotar o mesmo sinal que h , enquanto $|J_E|$ é invariavelmente positivo. Por consequência, há três resultados possíveis:

$$\begin{aligned}(vi) \quad h < 0 & \Rightarrow |J_E| > 0; \text{tr } J_E < 0 \\ (vii) \quad h > 0 & \Rightarrow |J_E| > 0; \text{tr } J_E > 0 \\ (viii) \quad h = 0 & \Rightarrow |J_E| > 0; \text{tr } J_E = 0\end{aligned}$$

Esses resultados estão associados respectivamente à flutuação amortecida, flutuação explosiva e flutuação uniforme. Em outras palavras, a Possibilidade *vi* implica um foco estável; a Possibilidade *vii*, um foco instável e a Possibilidade *viii*, um vórtice.

As conclusões da discussão precedente estão resumidas na Tabela 19.1 para facilitar inferências qualitativas dos sinais de $|J_E|$ e $\text{tr } J_E$. Três das características da tabela são especialmente dignas de nota. Em primeiro lugar, um $|J_E|$ negativo é vinculado exclusivamente ao tipo de equilíbrio de ponto de sela. Isso sugere que $|J_E| < 0$ é uma condição necessária e suficiente para um ponto de sela. Em segundo lugar, um valor zero para $\text{tr } J_E$ ocorre somente sob duas circunstâncias – quan-

do há um ponto de sela ou um vórtice. Todavia, essas duas circunstâncias são distinguíveis entre si pelo sinal de $|J_E|$. De acordo com isso, um $\text{tr } J_E$ igual a zero juntamente com um $|J_E|$ positivo é necessário e suficiente para um vórtice. Em terceiro lugar, conquanto um sinal negativo para $\text{tr } J_E$ seja necessário para a estabilidade dinâmica, ele *não* é suficiente por causa da possibilidade de um ponto de sela. Não obstante, quando um $\text{tr } J_E$ negativo é acompanhado por um $|J_E|$ positivo, temos realmente uma condição necessária e suficiente para a estabilidade dinâmica.

Caso	Sinal de $ J_E $	Sinal de $\text{tr } J_E$	Tipo de Equilíbrio
1. $(\text{tr } J_E)^2 > 4 J_E $	+	–	Nó estável
	+	+	Nó instável
	–	+, 0, –	Ponto de sela
2. $(\text{tr } J_E)^2 = 4 J_E $	+	–	Nó estável
	+	+	Nó instável
3. $(\text{tr } J_E)^2 < 4 J_E $	+	–	Foco estável
	+	+	Foco instável
	+	0	Vórtice

TABELA 19.1

Análise da estabilidade local de um sistema de equações diferenciais não-linear de duas variáveis

A discussão que leva ao resumo apresentado na Tabela 19.1 foi conduzida no contexto de uma aproximação linear de um sistema *não-linear*. Entretanto, é óbvio que o conteúdo daquela tabela também é aplicável à análise qualitativa de um sistema que, já de início, é *linear*. Neste último caso, os elementos da matriz jacobiana serão um conjunto de constantes dadas, de modo que não há necessidade de calculá-las no ponto de equilíbrio. Uma vez que não há processo de aproximação envolvido, as inferências de estabilidade já não serão mais de natureza “local”, mas terão uma validade global.

EXEMPLO 1 Analise a estabilidade local do sistema não-linear

$$\begin{aligned} x' &= f(x, y) = xy - 2 \\ y' &= g(x, y) = 2x - y \quad (x, y \geq 0) \end{aligned}$$

Em primeiro lugar, fazendo $x' = y' = 0$ e notando a não negatividade de x e y , encontramos um único ponto de equilíbrio E em $(\bar{x}, \bar{y}) = (1, 2)$. Então, tomando as derivadas parciais de x' e y' , e calculando-as em E , obtemos

$$J_E = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} = \begin{bmatrix} y & x \\ 2 & -1 \end{bmatrix}_{(1, 2)} = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$$

Visto que $|J_E| = -4$ é negativo, podemos concluir, de imediato, que o ponto de equilíbrio é localmente um ponto de sela.

Note que, enquanto a primeira linha da matriz jacobiana contém originalmente as variáveis y e x , a segunda linha não as contém. A razão para a diferença é que a segunda equação do sistema dado é originalmente linear e não requer nenhuma linearização.

EXEMPLO 2 Dado o sistema não-linear

$$\begin{aligned} x' &= x^2 - y \\ y' &= 1 - y \end{aligned}$$

podemos, estabelecendo $x' = y' = 0$, encontrar dois pontos de equilíbrio: $E_1 = (1, 1)$ e $E_2 = (-1, 1)$. Portanto, precisamos de duas linearizações separadas. Calculando o jacobiano $\begin{bmatrix} 2x & -1 \\ 0 & -1 \end{bmatrix}$ nos dois pontos de equilíbrios, um por vez, obtemos

$$J_{E1} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \quad \text{e} \quad J_{E2} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix}$$

O primeiro dos dois tem um determinante negativo; assim, $E_1 = (1, 1)$ é localmente um ponto de sela. Pelo segundo, constatamos que $|J_{E2}| = 2$ e $\text{tr } J_{E2} = -3$. Em consequência, pela Tabela 19.1, $E_2 = (-1, 1)$ é localmente um nó estável sob o Caso 1.

EXEMPLO 3 O sistema linear

$$\begin{aligned} x' &= x - y + 2 \\ y' &= x + y + 4 \end{aligned}$$

possui um equilíbrio estável? Para responder a essa pergunta qualitativa, podemos simplesmente focalizar as equações homogêneas e ignorar completamente as constantes 2 e 4. Como é de se esperar de um sistema linear, os elementos do jacobiano $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ são quatro constantes.

Considerando que seu determinante e traço são ambos iguais a 2, o equilíbrio é classificado como Caso 3 e é um foco instável. Note que chegamos a essa conclusão sem ter de calcular o ponto de equilíbrio. Note, também, que, neste caso, a conclusão é globalmente válida.

EXEMPLO 4 Analise a estabilidade local do modelo de Obst (19.50),

$$\begin{aligned} p' &= h(1 - \mu) \\ \mu' &= (p + q - m)\mu \end{aligned}$$

supondo que a taxa de expansão monetária m é exógena (nenhuma regra monetária é seguida). Segundo a Figura 19.4a, o equilíbrio desse modelo ocorre em $E = (\bar{p}, \bar{\mu}) = (m - q, 1)$. A matriz jacobiana calculada em E é

$$J_E = \begin{bmatrix} \frac{\partial p'}{\partial p} & \frac{\partial p'}{\partial \mu} \\ \frac{\partial \mu'}{\partial p} & \frac{\partial \mu'}{\partial \mu} \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ \mu & p + q - m \end{bmatrix}_{(m-q, 1)} = \begin{bmatrix} 0 & -h \\ 1 & 0 \end{bmatrix}$$

Uma vez que $|J_E| = h > 0$ e $\text{tr } J_E = 0$, a Tabela 19.1 indica que o equilíbrio é localmente um vórtice. Essa conclusão é consistente com a da análise do diagrama de fase na Seção 19.5.

EXEMPLO 5 Analise a estabilidade local do modelo de Obst, supondo que a regra monetária alternativa é a seguinte:

$$\begin{aligned} p' &= h(1 - \mu) && \text{[de (19.50)]} \\ \mu' &= [p + q - m(p')]\mu && \text{[de (19.58)]} \end{aligned}$$

Note que, uma vez que p' é uma função de μ , a função $m(p')$ é, no presente modelo, também uma função de μ . Assim, temos de aplicar a regra do produto para encontrar $\partial \mu' / \partial \mu$. No equilíbrio E , onde $p' = \mu' = 0$, temos $\bar{\mu} = 1$ e $\bar{p} = m(0) - q$. O jacobiano calculado em E é, portanto,

$$J_E = \begin{bmatrix} 0 & -h \\ \mu & p + q - m(p') - m'(p')(-h)\mu \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ 1 & m'(0)h \end{bmatrix}$$

onde $m'(0)$ é negativo por (19.57). De acordo com a Tabela 19.1, com $|J_E| = h > 0$ e $\text{tr } J_E = m'(0)h < 0$, podemos ter ou um foco estável ou um nó estável, dependendo das grandezas relativas de $(\text{tr } J_E)^2$ e $4|J_E|$. Para sermos específicos, quanto maior o valor absoluto da derivada $m'(0)$, maior será o valor absoluto de $\text{tr } J_E$ e mais provável será que $(\text{tr } J_E)^2$ exceda $4|J_E|$, para produzir um nó estável em vez de um foco estável. Esta conclusão novamente é consistente com o que aprendemos da análise do diagrama de fase.

EXERCÍCIO 19.6

1. Analise a estabilidade local de cada um dos seguintes sistemas não-lineares:

(a) $x' = e^x - 1$

$y' = ye^x$

(b) $x' = x + 2y$

$y' = x^2 + y$

(c) $x' = 1 - e^y$

$y' = 5x - y$

(d) $x' = x^3 + 3x^2 y + y$

$y' = x(1 + y^2)$

2. Use a Tabela 19.1 para determinar o tipo de equilíbrio que um sistema não-linear teria localmente, dado que:

(a) $f_x = 0$ $f_y > 0$ $g_x > 0$ e $g_y = 0$

(b) $f_x = 0$ $f_y < 0$ $g_x < 0$ e $g_y = 0$

(c) $f_x < 0$ $f_y > 0$ $g_x < 0$ e $g_y < 0$

Seus resultados são consistentes com suas respostas aos Exercícios 19.5–4 e 19.5–5?

3. Analise a estabilidade local do modelo de Obst supondo que seja seguida a regra monetária convencional.
4. Os dois sistemas seguintes possuem jacobianos de valor zero. Construa um diagrama de fase para cada um e deduza as localizações de todos os equilíbrios que existem:

(a) $x' = x + y$

$y' = -x - y$

(b) $x' = 0$

$y' = 0$



CAPÍTULO 20

Teoria do controle ótimo

No final do Capítulo 13, nos referimos à otimização dinâmica como um tipo de problema que não estávamos aptos a resolver porque ainda não tínhamos as ferramentas de análise dinâmica, tal como as equações diferenciais. Agora que já adquirimos essas ferramentas, por fim podemos experimentar um pouco da otimização dinâmica.

A abordagem clássica da otimização dinâmica é denominada *cálculo de variações*. Entretanto, em desenvolvimentos posteriores dessa metodologia, uma abordagem mais poderosa conhecida como *teoria do controle ótimo* suplantou, em grande parte, o cálculo de variações. Por essa razão, neste capítulo limitaremos nossa atenção à teoria do controle ótimo, explicando sua natureza básica, apresentando a importante ferramenta de solução denominada *princípio do máximo* e ilustrando sua utilização em alguns modelos econômicos elementares.[†]

20.1 A natureza do controle ótimo

Na otimização estática, a tarefa é encontrar um único valor para cada variável de escolha, tal que uma função objetivo enunciada seja maximizada ou minimizada, conforme o caso. Esse problema é desprovido de uma dimensão de tempo. Ao contrário, em um problema de otimização dinâmica, o tempo entra de modo explícito e destacado. Nesse tipo de problema, sempre teremos em mente um período de planejamento, digamos, a partir de um tempo inicial $t = 0$ até um tempo terminal $t = T$, e tentaremos encontrar o melhor curso de ação a ser adotado durante todo esse período. Assim, a solução para qualquer variável não tomará a forma de um valor único, mas de uma trajetória temporal completa.

Suponha que o problema seja maximizar lucros durante um período de tempo. Em qualquer instante de tempo t , temos de escolher o valor de alguma *variável de controle*, $u(t)$, que então afetará o valor de alguma *variável de estado*, $y(t)$, por meio do que denominamos uma *equação de movimento*. Por sua vez, $y(t)$ determinará o lucro $\pi(t)$. Visto que nosso objetivo é maximizar o lucro durante todo o período, a função objetivo deve tomar a forma de uma integral definida de π de $t = 0$ a $t = T$. Para ser completo, o problema também especifica o valor inicial da variável de estado y , $y(0)$, e o valor terminal de y , $y(T)$, ou, alternativamente, a faixa de valores permitidos que $y(T)$ pode assumir.

[†] Se quiser um tratamento mais completo da teoria do controle ótimo (bem como do "cálculo de variações") consulte *Elements of Dynamic Optimization*, de Alpha C. Chiang, McGraw-Hill, Nova York, 1992, agora publicado por Waveland Press, Inc., Prospect Heights, Illinois. Este capítulo utiliza muito do material publicado no livro citado.

Levando em conta o precedente, podemos enunciar o problema mais simples de controle ótimo como:

$$\begin{array}{ll}
 \text{Maximizar} & \int_0^T F(t, y, u) \, dt \\
 \text{Sujeita a} & \frac{dy}{dt} \equiv y' = f(t, y, u) \\
 \text{e} & \begin{array}{ll} y(0) = A & y(T) \text{ livre} \\ u(t) \in U & \text{para todo } t \in [0, T] \end{array}
 \end{array} \quad (20.1)$$

A primeira linha de (20.1), a função objetivo, é uma integral cujo integrando $F(t, y, u)$ estipula como a escolha da variável de controle u no instante t , juntamente com o y resultante no instante t , determina nosso objeto de maximização em t . A segunda linha é a equação de movimento para a variável de estado y . O que essa equação faz é proporcionar o mecanismo pelo qual a escolha que fizemos da variável de controle u pode ser traduzida para um padrão específico de movimento da variável de estado y . Normalmente, a ligação entre u e y pode ser descrita adequadamente por uma equação diferencial de primeira ordem $y' = f(t, y, u)$. Contudo, se acontecer de o padrão de variação da variável de estado requerer uma equação diferencial de segunda ordem, então devemos transformar essa equação em um par de equações diferenciais de primeira ordem. Nesse caso, será introduzida uma variável de estado adicional. Admite-se que o integrando F e a equação de movimento são ambos contínuos em todos os seus argumentos e possuem derivadas parciais de primeira ordem em relação à variável de estado y e à variável de tempo t , mas não necessariamente em relação à variável de controle u . Na terceira linha, indicamos que o estado inicial, o valor de y em $t=0$, é uma constante A , mas o estado terminal $y(T)$ fica sem restrições. Por fim, a quarta linha indica que as escolhas permissíveis de u estão limitadas a uma região de controle U . É claro que pode acontecer de $u(t)$ não ser restrita.

Ilustração: um modelo macroeconômico simples

Considere uma economia que produz resultado Y utilizando capital K e uma quantidade fixa de trabalho L , segundo a função produção

$$Y = Y(K, L)$$

Além disso, o produto resultante é utilizado ou para consumo C ou para investimento I . Se ignorarmos o problema da depreciação, então

$$I \equiv \frac{dK}{dt}$$

Em outras palavras, investimento é a variação no estoque de capital ao longo do tempo. Assim, também podemos escrever investimento como

$$I = Y - C = Y(K, L) - C = \frac{dK}{dt}$$

que nos dá uma equação diferencial de primeira ordem na variável K .

Se nosso objetivo for maximizar algum tipo de utilidade social durante um período de planejamento fixo, então o problema se torna

$$\begin{array}{ll}
 \text{Maximizar} & \int_0^T U(C) \, dt \\
 \text{sujeita a} & \frac{dK}{dt} = Y(K, L) - C \\
 \text{e} & K(0) = K_0 \quad K(T) = K_T
 \end{array} \quad (20.2)$$

onde K_0 e K_T são o valor inicial e o valor terminal (alvo) de K . Note que em (20.2) o estado terminal é um valor fixo, e não um valor livre, como em (20.1). Aqui, C funciona como a variável de controle e K é a variável de estado. O problema é escolher a trajetória de controle ótimo $C(t)$ de modo que seu impacto sobre o produto Y e o capital K , e a repercussão destes sobre C em si, juntos, maximizarão a utilidade agregada durante o período de planejamento.

Princípio do máximo de Pontryagin

A chave da teoria do controle ótimo é uma condição necessária de primeira ordem conhecida como o *princípio do máximo*.[†] O enunciado do princípio do máximo envolve uma abordagem semelhante à função de Lagrange e à variável do multiplicador de Lagrange. Para problemas de controle ótimo, estas são conhecidas como *função hamiltoniana* e *variável de coestado*, conceitos que desenvolveremos agora.

A hamiltoniana

Em (20.1), há três variáveis: o tempo t , a variável de estado y e a variável de controle u . Agora vamos introduzir uma nova variável, conhecida como variável de coestado e denotada por $\lambda(t)$. Assim como o multiplicador de Lagrange, a variável de coestado mede o preço sombra da variável de estado.

A variável de coestado é introduzida no problema do controle ótimo por meio de uma *função hamiltoniana* (ou abreviadamente, hamiltoniana). A hamiltoniana é definida como

$$H(t, y, u, \lambda) \equiv F(t, y, u) + \lambda(t) f(t, y, u) \quad (20.3)$$

onde H denota a hamiltoniana e é uma função de quatro variáveis: t , y , u e λ .

O princípio do máximo

O princípio do máximo – a principal ferramenta para resolver problemas de controle ótimo – tem esse nome porque, por ser uma condição necessária de primeira ordem, requer que escolhamos u de modo a maximizar a hamiltoniana H em todos os instantes.

Uma vez que, à parte a variável de controle, u , H envolve a variável de estado y e a variável de coestado λ , o enunciado do princípio do máximo também estipula como y e λ devem variar ao longo do tempo, por meio de uma *equação de movimento para a variável de estado* y (abreviadamente, *equação de estado*), bem como de uma *equação de movimento para a variável de coestado* λ (abreviadamente, *equação de coestado*). A equação de estado sempre vem como parte do enunciado do problema em si, como na segunda equação em (20.1). Mas, visto que (20.3) implica $\partial H / \partial \lambda = f(t, y, u)$, o princípio do máximo descreve a equação de estado

$$y' = f(t, y, u) \text{ como } y' = \frac{\partial H}{\partial \lambda} \quad (20.4)$$

Ao contrário, λ não aparece no enunciado do problema (20.1) e sua equação de movimento entra em cena exclusivamente como uma condição de otimização. A equação de coestado é

$$\lambda' \left(\equiv \frac{d\lambda}{dt} \right) = - \frac{\partial H}{\partial y} \quad (20.5)$$

Note que ambas as equações de movimento são enunciadas em termos das derivadas parciais de H , sugerindo uma certa simetria, mas há um sinal negativo ligado a $\partial H / \partial y$ em (20.5).

[†] O termo “princípio do máximo” é atribuído a L. S. Pontryagin e seus associados e freqüentemente é denominado princípio do máximo de Pontryagin. Veja *The Mathematical Theory of Optimal Control Processes*, de L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze e E. F. Mishchenko, Interscience, Nova York, 1962 (traduzido por K. N. Trilogoff).

As equações (20.4) e (20.5) constituem um sistema de duas equações diferenciais. Assim, precisamos de duas condições de fronteira para definir as duas constantes arbitrárias que surgirão no processo de solução. Se o estado inicial $y(0)$ e o estado terminal $y(T)$ forem ambos fixos, então essas especificações podem ser usadas para definir as constantes. Mas, se, como acontece no problema (20.1), o estado terminal não for fixo, então é preciso incluir o que denominamos uma *condição de transversalidade* como parte do princípio do máximo, para preencher a lacuna deixada pela condição de fronteira que está faltando.

Resumindo o precedente, podemos expressar os vários componentes do princípio do máximo para o problema (20.1) da seguinte maneira:

$$\begin{aligned}
 (i) \quad & H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) && \text{para todo } t \in [0, T] && (20.6) \\
 (ii) \quad & y' = \frac{\partial H}{\partial \lambda} && \text{(equação de estado)} \\
 (iii) \quad & \lambda' = -\frac{\partial H}{\partial y} && \text{(equação de coestado)} \\
 (iv) \quad & \lambda(T) = 0 && \text{(condição de transversalidade)}
 \end{aligned}$$

A Condição *i* em (20.6) afirma que, em cada instante t , o valor de $u(t)$, o controle ótimo, deve ser escolhido de modo a maximizar o valor da hamiltoniana em todos os valores admissíveis de $u(t)$. No caso em que a hamiltoniana for diferenciável em relação a u e der como resultado uma solução interior, a Condição *i* pode ser substituída por

$$\frac{\partial H}{\partial u} = 0$$

Contudo, se a região de controle for um conjunto fechado, então soluções de fronteira são possíveis e $\partial H / \partial u = 0$ podem não se aplicar. De fato, o princípio do máximo nem mesmo requer que a hamiltoniana seja diferenciável em relação a u .

As Condições *ii* e *iii* do princípio do máximo, $y' = \partial H / \partial \lambda$ e $\lambda' = -\partial H / \partial y$, nos dão duas equações de movimento, denominadas o sistema hamiltoniano para o problema dado. A Condição *iv*, $\lambda(T) = 0$, é a condição de transversalidade adequada apenas ao problema do estado terminal livre.

EXEMPLO 1

Para ilustrar a utilização do princípio do máximo, em primeiro lugar vamos considerar um exemplo não-econômico simples – o de encontrar a trajetória mais curta de um dado ponto A até uma reta dada. Na Figura 20.1, colocamos o ponto A no eixo vertical no plano ty e desenhemos a reta como uma reta vertical em $t = T$. São mostradas três trajetórias admissíveis (de um número infinito de trajetórias), cada uma com um comprimento diferente. O comprimento de qualquer trajetória é o agregado de pequenos segmentos de trajetória e cada um deles pode ser considerado como a hipotenusa (não desenhada) de um triângulo formado por pequenos movimentos dt e dy . Denotando a hipotenusa por dh , temos, pelo teorema de Pitágoras,

$$dh^2 = dt^2 + dy^2$$

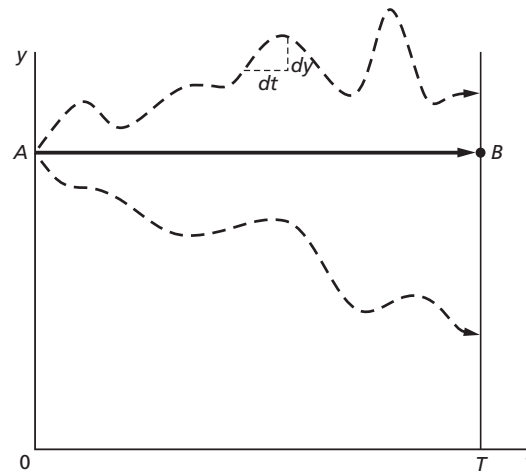
Dividindo ambos os lados por dt^2 e extraindo a raiz quadrada, obtemos

$$\frac{dh}{dt} = \left[1 + \left(\frac{dy}{dt} \right)^2 \right]^{1/2} = [1 + (y')^2]^{1/2} \quad (20.7)$$

Então, o comprimento total da trajetória pode ser encontrado integrando (20.7) em relação a t , de $t = 0$ a $t = T$. Se representarmos a variável de controle por $y' = u$, (20.7) pode ser expressa como

FIGURA 20.1

CHIANG / MAINWRIGHT



$$\frac{dh}{dt} = (1 + u^2)^{1/2} \quad (20.7')$$

É claro que minimizar a integral de (20.7') equivale a maximizar a *negativa* de (20.7'). Assim, o problema da trajetória mais curta é:

$$\begin{aligned} &\text{Maximizar} && \int_0^T -(1 + u^2)^{1/2} dt \\ &\text{sujeita a} && y' = u \\ &\text{e} && y(0) = A \quad y(T) \text{ livre} \end{aligned}$$

A hamiltoniana para o problema é, por (20.3),

$$H = -(1 + u^2)^{1/2} + \lambda u$$

Uma vez que H é diferenciável em u , e u não é restrita, a seguinte condição de primeira ordem pode ser utilizada para maximizar H :

$$\frac{\partial H}{\partial u} = -\frac{1}{2}(1 + u^2)^{-1/2}(2u) + \lambda = 0$$

$$\text{ou} \quad u(t) = \lambda(1 - \lambda^2)^{-1/2}$$

Verificando a condição de segunda ordem, constatamos que

$$\frac{\partial^2 H}{\partial u^2} = -(1 + u^2)^{-3/2} < 0$$

que comprova que a solução para $u(t)$ realmente maximiza a hamiltoniana. Visto que $u(t)$ é uma função de λ , precisamos de uma solução para a variável de coestado. Pelas condições de primeira ordem, a equação de movimento para a variável de coestado é

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

uma vez que H é independente de y . Assim, λ é uma constante. Para definir essa constante, podemos utilizar a condição de transversalidade $\lambda(T) = 0$. Visto que λ só pode assumir um único valor, que agora sabemos ser zero, na verdade temos $\lambda(t) = 0$ para todo t . Assim, podemos escrever

$$\lambda^*(t) = 0 \quad \text{para todo } t \in [0, T]$$

Deduz-se que o controle ótimo é

$$u^*(t) = \lambda^*[1 - (\lambda^*)^2]^{-1/2} = 0$$

Por fim, utilizando a equação de movimento para a variável de estado, constatamos que

$$\text{ou} \quad \begin{aligned} y' &= u = \\ y^*(t) &= c_0 \end{aligned} \quad (\text{uma constante})$$

Incorporando a condição inicial

$$y(0) = A$$

podemos concluir que $c_0 = A$, e escrever

$$y^*(t) = A \quad \text{para todo } t$$

Na Figura 20.1, essa trajetória é a reta AB . Verificamos que a trajetória mais curta é uma reta de inclinação zero.

EXEMPLO 2 Encontre a trajetória de controle ótimo que

$$\begin{aligned} &\text{Maximizará} && \int_0^1 (y - u^2) dt \\ &\text{sujeita a} && y' = u \\ &\text{e} && y(0) = 5 \quad y(1) \text{ livre} \end{aligned}$$

Este problema está no formato de (20.1), exceto que u não é restrita.

A hamiltoniana para esse problema,

$$H = y - u^2 + \lambda u$$

é côncava em u , e u não é restrita, portanto podemos maximizar H aplicando a condição de primeira ordem (também suficiente por causa da concavidade de H):

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

o que nos dá

$$u(t) = \frac{\lambda}{2} \quad \text{ou} \quad y' = \frac{\lambda}{2} \quad (20.8)$$

A equação de movimento para λ é

$$\lambda' = \frac{\partial H}{\partial y} = -1 \quad (20.8')$$

As duas últimas equações constituem o sistema de equações diferenciais para esse problema.

Podemos resolver primeiro para λ por integração direta de (20.8') para obter

$$\lambda(t) = c_1 - t \quad (c_1 \text{ arbitrária})$$

Além disso, pela condição de transversalidade em (20.6), devemos ter $\lambda(1) = 0$. Fazendo $t = 1$ na última equação, temos como resultado $c_1 = 1$. Assim, a trajetória de coestado ótima é

$$\lambda^*(t) = 1 - t$$

Segue-se que $y' = \frac{1}{2}(1 - t)$, por (20.8), e, por integração,

$$y(t) = \frac{1}{2}t - \frac{1}{4}t^2 + c_2 \quad (c_2 \text{ arbitrária})$$

A constante arbitrária pode ser definida utilizando-se a condição inicial $y(0) = 5$. Fazendo $t = 0$ na equação precedente, obtemos $5 = y(0) = c_2$. Assim, a trajetória ótima para a variável de estado é

$$y^*(t) = \frac{1}{2}t - \frac{1}{4}t^2 + 5$$

e a trajetória de controle ótima correspondente é

$$u^*(t) = \frac{1}{2}(1 - t)$$

EXEMPLO 3

Encontre a trajetória de controle ótima que

$$\begin{aligned} &\text{Maximizará} && \int_0^2 (2y - 3u) dt \\ &\text{sujeita a} && y' = y + u \\ &&& y(0) = 4 \quad y(2) \text{ livre} \\ &\text{e} && u(t) \in [0, 2] \end{aligned}$$

O fato de a variável de controle estar restrita ao conjunto fechado $[0, 2]$ dá origem à possibilidade de soluções de fronteira.

A função hamiltoniana

$$H = 2y - 3u + \lambda(y + u) = (2 + \lambda)y + (\lambda - 3)u$$

é linear em u . Se fizermos o gráfico de H em relação a u no plano uH , obteremos uma reta com inclinação $\partial H / \partial u = \lambda - 3$, que é positiva se $\lambda > 3$ (Reta 1), mas negativa se $\lambda < 3$ (Reta 2), como ilustrado na Figura 20.2. Se, em qualquer instante, λ exceder 3, então o máximo H ocorre na fronteira superior da região de controle e devemos escolher $u = 2$. Se, por outro lado, λ ficar abaixo de 3, então, para maximizar H , devemos escolher $u = 0$. Em suma, $u^*(t)$ depende de $\lambda(t)$, como se segue:

$$u^*(t) = \begin{cases} 2 \\ 0 \end{cases} \quad \text{e} \quad \lambda(t) \begin{cases} > \\ < \end{cases} 3 \quad (20.9)$$

Portanto, é fundamental encontrar $\lambda(t)$. Para fazer isso, começamos pela equação de coestado

$$\lambda' = -\frac{\partial H}{\partial y} = -2 - \lambda \quad \text{ou} \quad \lambda' + \lambda = -2$$

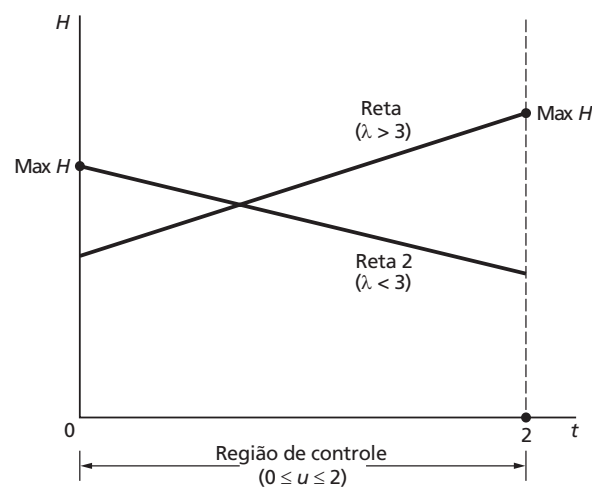


FIGURA 20.2

A solução geral dessa equação é

$$\lambda(t) = Ae^{-t} - 2 \quad [\text{por (15.5)}]$$

onde A é uma constante arbitrária. Utilizando a condição de transversalidade $\lambda(T) = \lambda(2) = 0$, constatamos que $A = 2e^2$. Assim, a solução definida para λ é

$$\lambda^*(t) = 2e^{2-t} - 2 \quad (20.10)$$

que é uma função decrescente de t , que diminui constantemente do valor inicial $\lambda^*(0) = 2e^2 - 2 = 12,778$ para um valor terminal $\lambda^*(2) = 2e^0 - 2 = 0$. Isso significa que λ^* tem de passar pelo ponto $\lambda = 3$ em algum instante crítico, τ , quando o u ótimo tiver de ser trocado de $u^* = 2$ para $u^* = 0$.

Para encontrar esse instante crítico, τ , fazemos $\lambda^*(\tau) = 3$ em (20.10):

$$3 = \lambda^*(\tau) = 2e^{2-\tau} - 2 \quad \text{ou} \quad e^{2-\tau} = \frac{5}{2} = 2,5$$

Tomando o logaritmo neperiano de ambos os lados, obtemos

$$\ln e^{2-\tau} = \ln 2,5 \quad \text{ou} \quad 2 - \tau = \ln 2,5$$

Assim,

$$\tau = 2 - \ln 2,5 = 1,084 \quad (\text{aprox.})$$

e o controle consiste em duas fases no intervalo de tempo $[0, 2]$:

$$\text{Fase 1: } u^*[0, \tau) = 2$$

$$\text{Fase 2: } u^*[\tau, 2] = 0$$

20.2 Condições terminais alternativas

O que acontece com o princípio do máximo quando a condição terminal for diferente da condição em (20.1)? Em (20.1), temos uma reta terminal vertical – com um tempo terminal fixo, mas estado terminal não-restrito, como ilustrado na Figura 20.1. O princípio do máximo para o problema de maximização requer que

$$(i) \quad H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) \quad \text{para todo } t \in [0, T]$$

$$(ii) \quad y' = \frac{\partial H}{\partial \lambda}$$

$$(iii) \quad \lambda' = -\frac{\partial H}{\partial y}$$

com a condição de transversalidade

$$(iv) \quad \lambda(T) = 0$$

Com condições terminais alternativas, as Condições *i*, *ii* e *iii* permanecerão as mesmas, mas a Condição *iv* (a condição de transversalidade) deve ser devidamente modificada.

Ponto terminal fixo

Se o ponto terminal for fixo de modo que a condição terminal é $y(T) = y_T$, sendo dados ambos T e y_T , então é a condição terminal em si que deve fornecer a informação para definir uma constante. Nesse caso, nenhuma condição de transversalidade é necessária.

Reta terminal horizontal

Suponha que o estado terminal é fixo em um determinado nível visado, y_T , mas o tempo terminal T é livre, de modo que temos a flexibilidade de alcançar o alvo a um ritmo rápido ou vagaroso. Então, temos uma reta horizontal terminal como a ilustrada na Figura 20.3a, que nos permite escolher entre T_1 , T_2 , T_3 ou outros tempos terminais para alcançar o nível visado, y . Para esse caso, a condição de transversalidade é uma restrição imposta à hamiltoniana (e não à variável de coestado) em $t = T$:

$$H_{t=T} = 0 \quad (20.11)$$

Reta terminal vertical truncada

Se tivermos um tempo terminal fixo T , e o estado terminal for livre, mas sujeito à condição $y_T \geq y_{\min}$, onde $y_{\min}$ denota um dado nível mínimo permissível de y , temos uma reta terminal vertical truncada, como a ilustrada na Figura 20.3b.

A condição de transversalidade para esse caso pode ser enunciada como a condição de folga complementar encontrada nas condições de Kuhn-Tucker:

$$\lambda(T) \geq 0 \quad y_T \geq y_{\min} \quad (y_T - y_{\min}) \lambda(T) = 0 \quad (20.12)$$

A abordagem prática para resolver esse tipo de problema é primeiro experimentar $\lambda(T) = 0$ como a condição de transversalidade e verificar se o y_T^* resultante satisfaz a restrição $y_T^* \geq y_{\min}$. Caso satisfaça, o problema está resolvido. Se não satisfizer, então trate o problema como o de um ponto terminal dado, tendo $y_{\min}$ como estado terminal.

Reta terminal horizontal truncada

Quando o estado terminal for fixo em y_T e o tempo terminal for livre, mas sujeito à restrição $T \leq T_{\max}$, onde $T_{\max}$ denota o último tempo permissível (um prazo final) para alcançar o y_T dado,

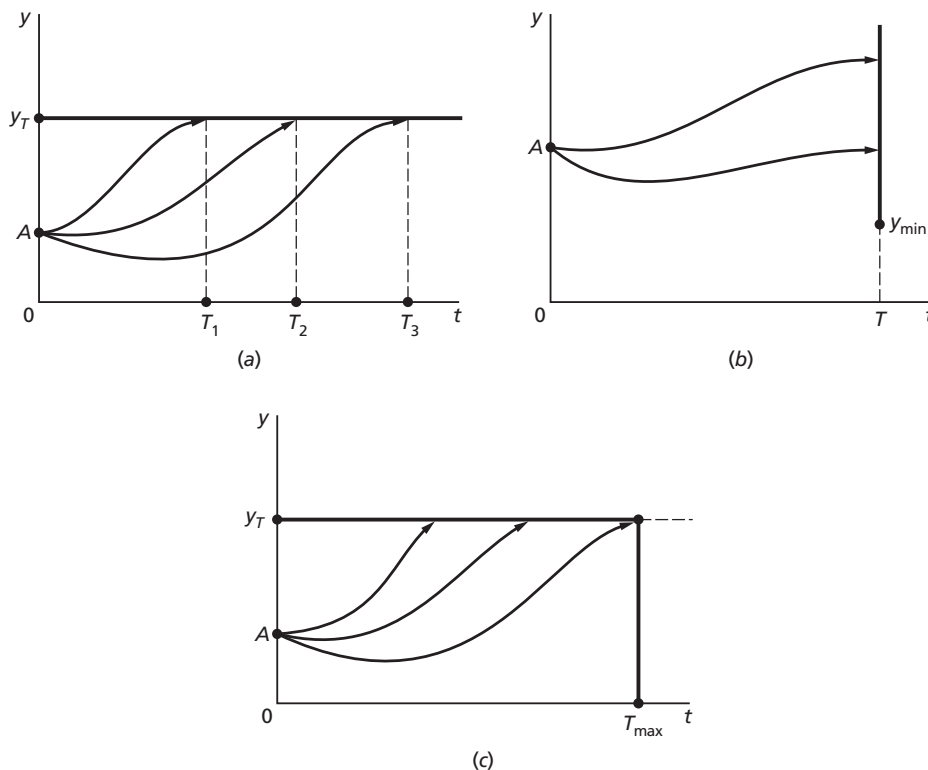


FIGURA 20.3

temos uma reta terminal horizontal truncada, como a ilustrada na Figura 20.3c. A condição de transversalidade se torna

$$H_{t=T_{\max}} \geq 0 \quad T \leq T_{\max} \quad (T - T_{\max}) H_{t=T_{\max}} = 0 \quad (20.13)$$

Mais uma vez, isso aparece no formato de uma condição de folga complementar.

A abordagem prática para resolver esse tipo de problema é experimentar primeiro $H_{t=T_{\max}} = 0$. Se o valor de solução resultante for $T^* \leq T_{\max}$, então o problema está resolvido e, assim, devemos tomar $T_{\max}$ como um tempo terminal fixo que, juntamente com o y_T dado, define um ponto final fixo, e resolver o problema como um problema de ponto final fixo.

EXEMPLO 1 No problema

$$\begin{array}{ll} \text{Maximizar} & \int_0^1 (y - u^2) dt \\ \text{sujeita a} & y' = u \\ \text{e} & y(0) = 2 \quad y(1) = a \end{array}$$

o ponto terminal é fixo, ainda que aqui tenhamos atribuído a $y(1)$ um valor paramétrico, e não um valor numérico.

A função hamiltoniana

$$H = y - u^2 + \lambda u$$

é côncava em u , portanto podemos fazer $\partial H / \partial u = 0$ para maximizar H :

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

Assim,

$$u = \frac{\lambda}{2}$$

o que mostra que, para resolver para $u(t)$, primeiro precisamos resolver para $\lambda(t)$.

As duas equações de movimento são

$$y' (= u) = \frac{\lambda}{2}$$

$$\lambda' \left(= -\frac{\partial H}{\partial y} \right) = -1$$

A integração direta da última equação dá como resultado

$$\lambda(t) = c_1 - t \quad (c_1 \text{ arbitrária})$$

o que implica que

$$y' = \frac{1}{2} c_1 - \frac{1}{2} t$$

Mais uma vez, por integração direta, constatamos que

$$y(t) = \frac{c_1}{2} t - \frac{1}{4} t^2 + c_2 \quad (c_2 \text{ arbitrária})$$

Para definir as duas constantes arbitrárias, utilizamos a condição inicial $y(0) = 2$ e a condição terminal $y(1) = a$. Fazendo $t = 0$ e $t = 1$, sucessivamente, na equação precedente, obtemos

$$2 = y(0) = c_2 \quad a = y(1) = \frac{c_1}{2} - \frac{1}{4} + c_2$$

$$\text{Assim, } c_2 = 2 \text{ e } c_1 = 2a - \frac{7}{2}.$$

Por conseguinte, podemos escrever as trajetórias ótimas deste problema como:

$$y^*(t) = (a - \frac{7}{4})t - \frac{1}{4}t^2 + 2$$

$$\lambda^*(t) = 2a - \frac{7}{2} - t$$

$$u^*(t) = a - \frac{7}{4} - \frac{1}{2}t$$

EXEMPLO 2 O problema

$$\begin{array}{ll} \text{Maximizar} & \int_0^T -(t^2 + u^2) dt \\ \text{sujeita a} & y' = u \\ \text{e} & y(0) = 4 \quad y(T) = 5 \quad T \text{ livre} \end{array}$$

exemplifica o caso da reta terminal horizontal cujo estado terminal é fixo, mas o tempo de chegada ao nível visado de y não é restrito. Na verdade, uma de nossas tarefas é resolver o valor ótimo de T .

Uma vez que a hamiltoniana

$$H = -t^2 - u^2 + \lambda u$$

é côncava em u , mais uma vez podemos maximizar H usando a condição de primeira ordem

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

que nos dá

$$u = \frac{\lambda}{2} \quad (20.14)$$

A concavidade de H torna desnecessário verificar a condição de segunda ordem, mas, se quisermos, é fácil verificar que $\partial^2 H / \partial u^2 = -2 < 0$, suficiente para um máximo de H .

A equação de movimento para λ é

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

o que implica que λ é uma constante. Mas ainda não podemos determinar seu exato valor neste ponto.

Voltando à equação de movimento para y ,

$$y' = u = \frac{\lambda}{2} \quad [\text{por (20.14)}]$$

podemos obter, por integração direta,

$$y(t) = \frac{\lambda}{2}t + c \quad (20.15)$$

Uma vez que $y(0) = 4$, vemos que $c = 4$. Além do mais, a condição de transversalidade (20.11) requer que

$$H_{t=T} = -T^2 - \frac{\lambda^2}{4} + \frac{\lambda^2}{2} = -T^2 + \frac{\lambda^2}{4} = 0 \quad [\text{por (20.14)}]$$

Resolvendo a equação precedente para T e tomando a raiz quadrada positiva, obtemos

$$T = \frac{\lambda}{2} \quad (20.16)$$

Uma vez que λ é constante, T também o é. Agora vamos tentar encontrar seu valor exato.

Aplicando a especificação de estado terminal $y(T) = 5$ a (20.15) e lembrando que $c = 4$, obtemos

$$y(T) = \frac{\lambda}{2} T + 4 = 5$$

Em vista de (20.16), a última equação pode ser reescrita como $T^2 = 1$. Assim, tomando a raiz quadrada, podemos determinar que o tempo de chegada ótimo é

$$T^* = 1 \quad (\text{raiz negativa inaceitável})$$

Então, podemos deduzir imediatamente que

$$\lambda^*(t) = 2T^* = 2 \quad [\text{por (20.16)}]$$

$$u^*(t) = \frac{\lambda}{2} = 1 \quad [\text{por (20.14)}]$$

$$y^*(t) = t + 4 \quad [\text{por (20.15)}]$$

O último resultado mostra que, neste exemplo, a trajetória ótima y é uma reta que vai do ponto inicial dado até a reta terminal horizontal.

EXERCÍCIO 20.2

Encontre as trajetórias ótimas das variáveis de controle, de estado e de coestado que

1. Maximizarão $\int_0^1 (y - u^2) dt$
 sujeita a $y' = u$
 e $y(0) = 2$ $y(1)$ livre
2. Maximizarão $\int_0^8 6y dt$
 sujeita a $y' = y + u$
 e $y(0) = 10$ $y(8)$ livre
 e $u(t) \in [0, 2]$
3. Maximizarão $\int_0^T -(au + bu^2) dt$
 sujeita a $y' = y - u$
 e $y(0) = y_0$ $y(t)$ livre
4. Maximizarão $\int_0^T (yu - u^2 - y^2) dt$
 sujeita a $y' = u$
 e $y(0) = y_0$ $y(t)$ livre
5. Maximizarão $\int_0^{20} -\frac{1}{2}u^2 dt$
 sujeita a $y' = u$
 e $y(0) = 10$ $y(20) = 0$
6. Maximizarão $\int_0^4 3y dt$
 sujeita a $y' = y + u$
 e $y(0) = 5$ $y(4) \geq 300$
 e $0 \leq u(t) \leq 2$

7. Maximizarão	$\int_0^1 -u^2 dt$
sujeita a	$y' = y + u$
e	$y(0) = 1 \quad y(1) = 0$
8. Maximizarão	$\int_1^2 (y + ut - u^2) dt$
sujeita a	$y' = u$
e	$y(1) = 3 \quad y(2) = 4$
9. Maximizarão	$\int_0^2 (2y - 3u - au^2) dt$
sujeita a	$y' = u + y$
e	$y(0) = 5 \quad y(2) \text{ livre}$

20.3 Problemas autônomos

Na estrutura geral do problema de controle, a variável t pode entrar diretamente na função objetivo e na equação de estado. A especificação geral

$$\begin{array}{ll} \text{Maximizar} & \int_0^T F(t, y, u) dt \\ \text{sujeita a} & y' = f(t, y, u) \\ \text{e} & \text{condições de fronteira} \end{array}$$

onde t entra explicitamente em F e f , significa que a data tem importância. Isto é, o valor gerado pela atividade $u(t)$ não depende apenas do nível, mas também de exatamente quando essa atividade ocorre.

Problemas nos quais t está ausente da função objetivo e da equação de estado, tal como

$$\begin{array}{ll} \text{Maximizar} & \int_0^T F(y, u) dt \\ \text{sujeita a} & y' = f(y, u) \\ \text{e} & \text{condições de fronteira} \end{array}$$

são denominados *problemas autônomos*. Nesses problemas, uma vez que a hamiltoniana

$$H = F(y, u) + \lambda f(y, u)$$

não contém t como um argumento, as equações de movimento são mais fáceis de resolver; além do mais, elas se prestam à utilização da análise de diagrama de fase.

E ainda em outros casos – em um problema que, não fosse por isso, seria autônomo –, o tempo t entra em cena como parte do fator de desconto $e^{-\rho t}$, mas em nenhum outro lugar, de modo que a função objetivo toma a forma de

$$\int_0^T G(y, u) e^{-\rho t} dt$$

Em termos estritos, esse problema não é autônomo. Todavia, é fácil converter o problema em autônomo empregando a denominada *hamiltoniana de valor corrente*, definida como:

$$H_c \equiv H e^{\rho t} = G(y, u) + \mu f(y, u) \quad (20.17)$$

onde

$$\mu \equiv \lambda e^{\rho t} \quad (20.18)$$

é o *multiplicador de Lagrange de valor corrente*. Focalizando o valor corrente (não descontado), podemos eliminar t da hamiltoniana original.

Usando H_c em lugar de H , devemos revisar o princípio do máximo para:

$$\begin{aligned}
 (i) \quad & H_c(y, u^*, \mu) \geq H_c(y, u, \mu) \quad \text{para todo } t \in [0, T] \\
 (ii) \quad & y' = \frac{\partial H_c}{\partial \mu} \\
 (iii) \quad & \mu' = -\frac{\partial H_c}{\partial y} + r\mu \quad (20.19) \\
 (iv) \quad & \mu(T) = 0 \quad (\text{para reta terminal vertical}) \\
 & \text{ou } [H_c]_{t=T} = 0 \quad (\text{para reta terminal horizontal})
 \end{aligned}$$

20.4 Aplicações econômicas

Maximização do tempo de vida da utilidade

Suponha que um consumidor tenha uma função utilidade $U(C(t))$, onde $C(t)$ é o consumo no instante t . A função utilidade do consumidor é côncava e tem as seguintes propriedades:

$$U' > 0 \quad U'' < 0$$

O consumidor também é dotado de um estoque inicial de riqueza, ou capital, K_0 , com fluxo de receita obtido do estoque de capital de acordo com o seguinte:

$$Y = rK$$

onde r é a taxa de juros do mercado. O consumidor utiliza a renda para comprar C . Além disso, ele pode consumir o estoque de capital. Qualquer receita não consumida é adicionada ao estoque de capital como investimento. Assim,

$$K' \equiv I = Y - C = rK - C$$

O problema de maximização do tempo de vida da utilidade do consumidor é

$$\begin{aligned}
 & \text{Maximizar} && \int_0^T U(C(t)) e^{-\delta t} dt \\
 & \text{sujeita a} && K' = rK(t) - C(t) \\
 & \text{e} && K(0) = K_0 \quad K(T) \geq 0
 \end{aligned}$$

onde δ é a taxa pessoal de preferência de tempo do consumidor ($\delta \geq 0$). Supomos que $C(t) > 0$ e $K(t) > 0$ para todo t .

A hamiltoniana é

$$H = U(C(t)) e^{-\delta t} + \lambda(t) [rK(t) - C(t)]$$

onde C é a variável de controle e K é a variável de estado. Uma vez que $U(C)$ é côncava e a restrição é linear em C , sabemos que a hamiltoniana é côncava e a maximização de H pode ser conseguida simplesmente fazendo $\partial H / \partial C = 0$. Assim, temos

$$\frac{\partial H}{\partial C} = U'(C) e^{-\delta t} - \lambda = 0 \quad (20.20)$$

$$K' = rK(t) - C(t) \quad (20.20')$$

$$\lambda' = -\frac{\partial H}{\partial K} = -r\lambda \quad (20.20'')$$

A equação (20.20) afirma que a utilidade marginal descontada deve ser igualada ao preço sombra presente de uma unidade adicional de capital. Diferenciando (20.20) em relação a t , obtemos

$$U''(C) C' e^{-\delta t} - \delta U'(C) e^{-\delta t} = \lambda' \quad (20.21)$$

Em vista de (20.20) e (20.20''), temos

$$\lambda' = -r\lambda = -rU'(C) e^{-\delta t}$$

que pode ser substituída em (20.21) para dar como resultado

$$U''(C) C' e^{-\delta t} - \delta U'(C) e^{-\delta t} = -rU'(C) e^{-\delta t}$$

ou, após simplificar o fator comum $e^{-\delta t}$ e rearranjar,

$$\frac{-U''(C(t))}{U'(C(t))} C'(t) = r - \delta$$

Uma vez que $U' > 0$ e $U'' < 0$, o sinal da derivada $C'(t)$ tem de ser o mesmo que o de $(r - \delta)$. Por conseguinte, se $r > \delta$, o consumo ótimo aumentará ao longo do tempo; se $r < \delta$, o consumo ótimo diminuirá ao longo do tempo.

Resolvendo (20.20'') diretamente, obtemos

$$\lambda(t) = \lambda_0 e^{-rt}$$

onde $\lambda_0 > 0$ é a constante de integração. Combinando essa expressão com (20.20), temos

$$U'(C(t)) = \lambda e^{\delta t} = \lambda_0 e^{(\delta-r)t}$$

que mostra que a utilidade marginal de consumo decrescerá otimamente ao longo do tempo se $r > \delta$, mas aumentará com o tempo se $r < \delta$.

Visto que a condição terminal $K(T) \geq 0$ identifica o presente problema como um problema de reta terminal vertical truncada, a condição de transversalidade adequada é, por (20.12),

$$\lambda(T) \geq 0 \quad K(T) \geq 0 \quad K(T)\lambda(T) = 0$$

A condição fundamental é a estipulação de folga complementar, significando que ou o estoque de capital K deve estar esgotado na data terminal, ou o preço sombra do capital λ deve cair a zero na data terminal. Por hipótese, $U'(C) > 0$, a utilidade marginal nunca pode ser zero. Portanto, o valor marginal do capital não pode ser zero. Isso implica que, neste modelo, o estoque de capital deve estar esgotado otimamente na data terminal T .

Recurso exaurível

Seja $s(t)$ o estoque de um recurso exaurível e $q(t)$ a taxa de extração em qualquer instante t , de modo que

$$s' = -q$$

O recurso extraído produz um bem de consumo final c tal que

$$c = c(q) \quad \text{onde} \quad c' > 0, c'' < 0 \quad (20.22)$$

O bem de consumo é o único argumento na função utilidade de um consumidor representativo e tem as seguintes propriedades:

$$U = U(c) \quad \text{onde} \quad U' > 0, U'' < 0 \quad (20.22')$$

O consumidor deseja maximizar a função utilidade dentro de um intervalo dado $[0, T]$. Uma vez que c é uma função de q , a taxa de extração, q funcionará como a variável de controle. Para simplificar, ignoramos a questão do desconto ao longo do tempo. Então, o problema dinâmico é escolher a taxa de extração ótima que maximizará a função utilidade sujeita somente à restrição de não-negatividade da variável de estado $s(t)$, o estoque do recurso exaurível. A formulação é

$$\begin{array}{ll} \text{Maximizar} & \int_0^T U(c(q)) \, dt \\ \text{sujeita a} & s' = -q \\ \text{e} & s(0) = s_0 \quad s(T) \geq 0 \end{array} \quad (20.23)$$

onde s_0 e T são dados.

A hamiltoniana para o problema é

$$H = U(c(q)) - \lambda q$$

Uma vez que H é côncava em q por especificações do modelo na função $U(c(q))$, podemos maximizar H fazendo $\partial H / \partial q = 0$:

$$\frac{\partial H}{\partial q} = U'(c(q)) c'(q) - \lambda = 0 \quad (20.24)$$

A concavidade de H nos assegura que (20.24) maximiza H , mas é fácil verificar a condição de segunda ordem e confirmar que $\partial^2 H / \partial q^2$ é negativa.

O princípio do máximo estipula que

$$\lambda' = -\frac{\partial H}{\partial s} = 0$$

o que implica que

$$\lambda(t) = c_0 \quad \text{uma constante} \quad (20.25)$$

Para determinar c_0 , recorreremos às condições de transversalidade. Uma vez que o modelo especifica $K(T) \geq 0$, ele tem uma reta terminal vertical truncada, de modo que (20.12) se aplica:

$$\lambda(T) \geq 0 \quad s(T) \geq 0 \quad s(T)\lambda(T) = 0$$

Em aplicações práticas, a etapa inicial é experimentar $\lambda(T) = 0$, resolver para q , e verificar se a solução funcionará. Visto que $\lambda(T)$ é uma constante, experimentar $\lambda(T) = 0$ implica $\lambda(t) = 0$ para todo t , e $\partial H / \partial q$ em (20.24) se reduz a

$$U'(c) c'(q) = 0$$

que (em princípio) pode ser resolvida para q . Uma vez que t não é um argumento explícito de U ou c , a trajetória de solução para q é constante ao longo do tempo:

$$q^*(t) = q^*$$

Agora, verificamos se q^* satisfaz a restrição $s(T) \geq 0$. Se q^* for uma constante, então a equação de movimento

$$s' = -q$$

pode ser imediatamente integrada, dando como resultado

$$s(t) = -qt + c_1 \quad [c_1 = \text{constante de integração}]$$

Utilizando a condição inicial $s(0) = s_0$, obtemos como resultado uma solução para a constante de integração

$$c_1 = s_0$$

e a trajetória de estado ótima é

$$s(t) = s_0 - q^*t \quad (20.26)$$

Sem especificar as formas funcionais para U e c , nenhuma solução numérica pode ser encontrada para q^* . Contudo, pelas condições de transversalidade, podemos concluir que, se $s(T) \geq 0$, então q^* , como obtida na solução, é aceitável. Mas, se $s(T) < 0$ para a q^* dada, então a taxa de extração é muito alta e precisamos achar uma solução diferente. Uma vez que a solução experimental $\lambda(T) = 0$ falhou, agora tomamos a alternativa de $\lambda(T) > 0$. Porém, mesmo nesse caso, λ ainda é uma constante por (20.25). E (20.24) ainda pode (em princípio) resultar em um valor de solução q_2^* constante, embora diferente. Deduz-se que (20.26) permanece válida. Mas, desta vez, com $\lambda(T) > 0$, a condição de transversalidade (20.12) determina que $s(T) = 0$ ou, em vista de (20.26),

$$s_0 - q_2^*T = 0$$

Assim, podemos escrever a taxa ótima de extração (constante) revisada como

$$q_2^* = \frac{s_0}{T}$$

Esse novo valor de solução deve representar uma taxa de extração mais baixa que não violaria a condição de fronteira $s(T) \geq 0$.

EXERCÍCIO 20.4

- Maximizar $\int_0^T (K - \alpha K^2 - I^2) dt$ ($\alpha > 0$)
sujeita a $K' = I - \delta K$ ($\delta > 0$)
e $K(0) = K_0$ $K(T)$ livre
- Resolva o seguinte problema de recurso exaurível para a trajetória de extração ótima:
Maximizar $\int_0^T \ln(q) e^{-\delta t} dt$
sujeita a $s' = -q$
e $s(0) = s_0$ $s(t) \geq 0$

20.5 Horizonte de tempo infinito

Nesta seção, apresentamos o problema da otimização dinâmica ao longo de um período de planejamento infinito. Modelos de horizonte de tempo infinito tendem a introduzir complexidades no que diz respeito às condições de transversalidade e trajetórias temporais ótimas, que são diferentes das que desenvolvemos antes. Em vez de atacar esses assuntos aqui, ilustraremos a metodologia desses modelos com uma versão do *modelo neoclássico de crescimento ótimo*.

Modelo neoclássico de crescimento ótimo

A função produção neoclássica padrão expressa a produção Y como uma função de dois insumos: trabalho L e capital K . Sua forma geral é

$$Y = Y(K, L)$$

onde $Y(K, L)$ é uma função linearmente homogênea com as seguintes propriedades:

$$Y_L > 0 \quad Y_K > 0 \quad Y_{LL} < 0 \quad Y_{KK} < 0$$

Reescrevendo a função produção em termos *per capita*, temos

$$y = \phi(k) \quad \text{com } \phi'(k) > 0 \quad \text{e} \quad \phi''(k) < 0$$

onde $y = Y/L$ e $k = K/L$. A produção total Y é alocada ao consumo C ou ao investimento bruto I . Seja δ a taxa de depreciação do estoque de capital K . Então, o investimento líquido ou as variações no estoque de capital podem ser escritos como

$$K' = I - \delta K = Y - C - \delta K$$

Denotando consumo *per capita* por $c \equiv C/L$, podemos escrever

$$\frac{1}{L} K' = y - c - \delta k \quad (20.27)$$

O lado direito de (20.27) está em termos *per capita*, mas o lado esquerdo, não. Para unificar, notamos que

$$K' = \frac{dk}{dt} = \frac{d}{dt}(kL) = k \frac{dL}{dt} + L \frac{dk}{dt} \quad (20.28)$$

Se a taxa de crescimento da população for [†]

$$\frac{dL/dt}{L} = n \quad \text{de modo que} \quad \frac{dL}{dt} = nL$$

então (20.28) se torna

$$K' = knL + Lk' \quad \text{ou} \quad \frac{1}{L} K' = kn + k'$$

Substituir essa expressão em (20.27) transforma a última em uma equação inteiramente em termos *per capita*:

$$k' = y - c - (n + \delta)k = \phi(k) - c - (n + \delta)k \quad (20.27')$$

Seja $U(c)$ a função bem-estar social (expressa em termos *per capita*), onde

[†] Neste modelo, admitimos que a força de trabalho e a população são uma só e a mesma.

$$U'(c) > 0 \quad \text{e} \quad U''(c) < 0$$

e, para eliminar soluções de vértice, também supomos que

$$U'(c) \rightarrow \infty \quad \text{quando } c \rightarrow 0$$

$$\text{e} \quad U'(c) \rightarrow 0 \quad \text{quando } c \rightarrow \infty$$

Se ρ denotar a taxa de desconto social e a população inicial for normalizada para uma unidade, a função objetivo pode ser expressa como

$$\begin{aligned} V &= \int_0^\infty U(c) e^{-\rho t} L_0 e^{nt} dt = \int_0^\infty U(c) e^{-(\rho-n)t} dt \\ &= \int_0^\infty U(c) e^{-rt} dt \quad \text{onde } r = \rho - n \end{aligned}$$

Nesta versão do modelo neoclássico de crescimento ótimo, a utilidade é ponderada por uma população que cresce continuamente a uma taxa n . Todavia, se $r = \rho - n > 0$, então, em termos matemáticos, o modelo não é diferente do modelo sem ponderações da população, mas com uma taxa positiva de desconto r .

O problema do crescimento ótimo agora pode ser enunciado como

$$\begin{aligned} &\text{Maximizar} && \int_0^\infty U(c) e^{-rt} dt && (20.29) \\ &\text{sujeita a} && k' = \phi(k) - c - (n + \delta)k \\ & && k(0) = k_0 \\ &\text{e} && 0 \leq c(t) \leq \phi(k) \end{aligned}$$

onde k é a variável de estado e c é a variável de controle.

A hamiltoniana para o problema é

$$H = U(c) e^{-rt} + \lambda [\phi(k) - c - (n + \delta)k]$$

Visto que H é côncava em c , o máximo de H corresponde a uma solução interior na região de controle $[0 < c < f(k)]$ e, por conseguinte, podemos encontrar o máximo de H por

$$\frac{\partial H}{\partial c} = U'(c) e^{-rt} - \lambda = 0$$

$$\text{ou} \quad U'(c) = \lambda e^{rt} \quad (20.30)$$

A interpretação econômica de (20.30) é que, ao longo da trajetória ótima, a utilidade marginal de consumo *per capita* deve ser igual ao preço sombra do capital (λ) ponderado por e^{rt} . Verificando as condições de segunda ordem, encontramos

$$\frac{\partial^2 H}{\partial c^2} = U''(c) e^{-rt} < 0$$

Por conseguinte, a hamiltoniana é maximizada.

Pelo princípio do máximo, temos duas equações de movimento

$$k' = \frac{\partial H}{\partial \lambda} = \phi(k) - c - (n + \delta)k$$

$$\text{e} \quad \lambda' = -\frac{\partial H}{\partial k} = -\lambda [\phi'(k) - (n + \delta)]$$

Em princípio, as duas equações de movimento combinadas com $U'(c) = \lambda e^{rt}$ devem definir uma solução para c, k, λ . Contudo, nesse nível de generalidade não podemos fazer mais do que executar uma análise qualitativa do modelo. Qualquer coisa a mais exigiria formas específicas de ambas as funções utilidade e produção.

A hamiltoniana de valor corrente

Uma vez que o modelo precedente é um exemplo de problema autônomo (t não é um argumento isolado na função utilidade ou equação de estado, mas aparece somente no fator de desconto), podemos usar a hamiltoniana de valor corrente escrita como

$$H_c = H e^{rt} = U(c) + \mu [\phi(k) - c - (n + \delta)k] \quad [\text{veja (20.17)}]$$

onde $\mu = \lambda e^{rt}$.

O princípio do máximo exige

$$\frac{\partial H_c}{\partial c} = U'(c) - \mu = 0 \quad \text{ou} \quad \mu = U'(c) \quad (20.31)$$

$$k' = \frac{\partial H_c}{\partial \mu} = \phi(k) - c - (n + \delta)k \quad (20.31')$$

$$\begin{aligned} \mu' &= \frac{\partial H_c}{\partial k} + r\mu = -\mu[\phi'(k) - (n + \delta)] + r\mu \\ &= -\mu[\phi'(k) - (n + \delta + r)] \end{aligned} \quad (20.31'')$$

As equações (20.31') e (20.31'') constituem um sistema de equações diferenciais autônomo, o que possibilita uma análise qualitativa por diagrama de fase.

Construindo um diagrama de fase

As variáveis nas equações diferenciais (20.31') e (20.31'') são k e μ . Visto que (20.31) envolve uma função de c , a saber, $U'(c)$, em vez da simples função c em si, seria mais simples construir um diagrama de fase no espaço kc , e não no espaço $k\mu$. Para fazer isso, tentaremos eliminar μ . Uma vez que $\mu = U'(c)$, por (20.31), a diferenciação em relação a t nos dá

$$\mu' = U''(c) c'$$

Substituindo essas expressões para μ e μ' em (20.31''), temos como resultado

$$c' = -\frac{U'(c)}{U''(c)} [\phi'(k) - (n + \delta + r)]$$

que é uma equação diferencial em c . Agora temos o sistema de equações diferenciais autônomo

$$k' = \phi(k) - c - (n + \delta)k \quad (20.31')$$

e

$$c' = -\frac{U'(c)}{U''(c)} [\phi'(k) - (n + \delta + r)] \quad (20.32)$$

Para construir o diagrama de fase no espaço kc , primeiro desenhemos as curvas $k' = 0$ e $c' = 0$, que são definidas por

$$c = \phi(k) - (n + \delta)k \quad (k' = 0) \quad (20.33)$$

e

$$\phi'(k) = n + \delta + r \quad (c' = 0) \quad (20.34)$$

Essas duas curvas estão ilustradas na Figura 20.4. A equação para a curva $k' = 0$, (20.33), tem a mesma estrutura da equação fundamental do modelo de crescimento de Solow, (15.30). Assim, a curva $k' = 0$ tem a mesma forma geral da curva na Figura 15.5b. Por outro lado, a representação gráfica da curva $c' = 0$ é uma reta vertical porque, dadas as especificações do modelo $\phi'(k) > 0$ e $\phi''(k) < 0$, $\phi(k)$ está associada a uma curva côncava de inclinação ascendente, com uma inclinação diferente em cada ponto da curva, de modo que somente um único valor de k pode satisfazer (20.34). A interseção das duas curvas no ponto E determina os valores de equilíbrio intertemporal de k e c porque, no ponto E , nem k nem c mudarão de valor ao longo do tempo, resultando em um *estado estacionário*. Poderíamos denominar esses valores $\bar{k}$ e $\bar{c}$ para valores de equilíbrio intertemporal, mas, em vez disso, nós os denominaremos k^* e c^* , porque eles também representam os valores de equilíbrio para o crescimento ótimo.

Análise do diagrama de fase

A interseção do ponto E na Figura 20.4 nos dá um estado estacionário único. Mas o que acontece se estivermos inicialmente em algum outro ponto que não o ponto E ? Voltando a nosso sistema de equações diferenciais de primeira ordem (20.31') e (20.32), podemos deduzir que

$$\frac{\partial k'}{\partial c} = -1 < 0 \quad \text{e} \quad \frac{\partial c'}{\partial k} = -\frac{U'(c)}{U''(c)} \phi''(k) < 0$$

Uma vez que $\partial k' / \partial c < 0$, todos os pontos abaixo da curva $k' = 0$ são caracterizados por $k' > 0$ e todos os pontos acima da curva, por $k' < 0$. De modo semelhante, visto que $\partial c' / \partial k < 0$, todos os pontos à esquerda da reta $c' = 0$ são caracterizados por $c' > 0$ e todos os pontos à direita da reta, por $c' < 0$. Assim, a curva $k' = 0$ e a reta $c' = 0$ dividem o espaço de fase em quatro regiões, cada uma com seus próprios e distintos pares de sinais de c' e k' . Isso é representado na Figura 20.5 pelas setas direcionais em ângulo reto que aparecem em cada região.

As linhas de fluxo que seguem as retas direcionais em cada região nos dizem que o estado estacionário no ponto E é um ponto de sela. Se tivermos um ponto inicial que esteja sobre um dos dois ramos estáveis do ponto de sela, a dinâmica do sistema nos levará ao ponto E . Mas qualquer ponto inicial que não esteja sobre um ramo estável ou nos fará contornar o ponto E , sem nunca alcançá-lo, ou nos afastará constantemente dele. Se seguirmos as linhas de fluxo dessas últimas instâncias, terminaremos, no seu devido tempo (quando $t \rightarrow \infty$), ou com $k = 0$ (esgotamento de capital) ou com $c = 0$ (consumo *per capita* minguando até zero) – ambas situações inaceitáveis do ponto de vista econômico. Assim, a única alternativa viável é escolher um par (k, c) que posicione

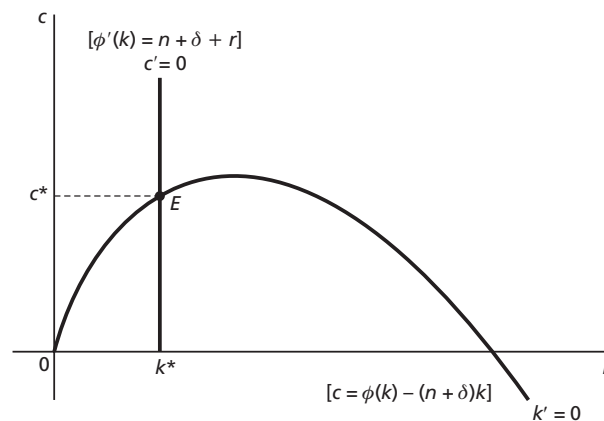
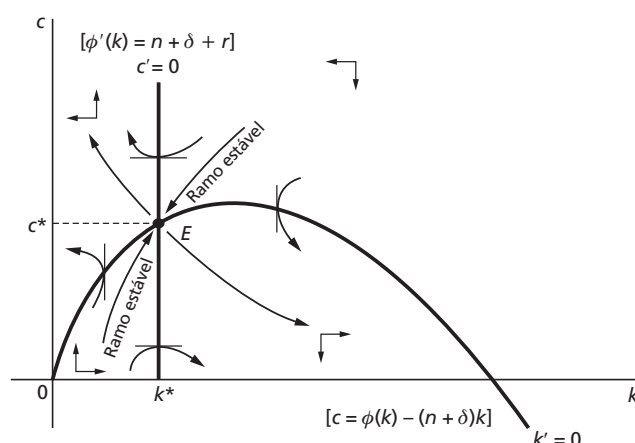


FIGURA 20.4

FIGURA 20.5



nossa economia sobre um ramo estável – uma “estrada de tijolos amarelos”, por assim dizer – que nos levará ao estado estacionário em E . Ainda não falamos explicitamente sobre a condição de transversalidade, mas, se tivéssemos falado, ela nos teria guiado ao estado estacionário em E , no qual o consumo *per capita* pode ser mantido em um nível constante para sempre.

20.6 Limitações da análise dinâmica

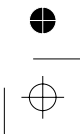
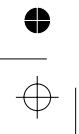
A análise estática apresentada na Parte 2 deste volume tratava apenas da questão de qual seria a posição de equilíbrio sob certas condições dadas de um modelo. A questão principal era: quais valores das variáveis que, *se atingidos*, tenderiam a se perpetuar? Mas a *atingibilidade* da posição de equilíbrio era dada como certa. Quando entramos no reino da estática comparativa, na Parte 3, a questão central passou a um problema mais interessante: como a posição de equilíbrio se deslocaria em resposta a certas variações em um parâmetro? Mas o aspecto da atingibilidade mais uma vez foi posto de lado. Somente quando chegamos à análise dinâmica, na Parte 5, encaramos de frente a questão da atingibilidade. Agora perguntamos especificamente: se estivéssemos, de início, longe de uma posição de equilíbrio – digamos, por causa de uma recente variação desequilibradora de um parâmetro –, as várias forças no modelo tenderiam a nos conduzir rumo a uma nova posição de equilíbrio? Além do mais, em uma análise dinâmica também aprendemos o caráter específico da trajetória (se estacionária, flutuante ou oscilatória) que a variável seguirá em seu caminho para o equilíbrio (se for o caso). Portanto, o significado da análise dinâmica deve ser evidente por si só.

Contudo, para concluir a discussão da análise dinâmica, também devemos tomar conhecimento de suas limitações. Uma razão é que, para podermos viabilizar a análise, muitas vezes os modelos dinâmicos são formulados em termos de equações lineares. Conquanto isso possa resultar em simplicidade, em muitos casos a premissa de linearidade acarretará um considerável sacrifício do realismo. Uma vez que uma trajetória que nasce de um modelo linear, nem sempre pode ser aproximada à de sua contraparte não-linear. Como vimos no exemplo do teto de preço na Seção 17.6, por exemplo, é preciso tomar cuidado na interpretação e aplicação dos resultados de modelos dinâmicos lineares. Nesse caso, entretanto, a abordagem gráfico-qualitativa pode prestar um serviço extremamente valioso porque, sob condições razoavelmente gerais, ela pode nos permitir incorporar a não-linearidade a um modelo sem acrescentar complexidade à análise.

Uma outra desvantagem encontrada com frequência em modelos econômicos dinâmicos é a utilização de coeficientes constantes em equações diferenciais ou de diferenças. Considerando que o papel principal dos coeficientes é especificar os parâmetros do modelo, a constância de coeficientes – mais uma vez suposta em favor da viabilidade matemática – serve, em essência, para “congelar” o ambiente econômico do problema sob investigação. Em outras palavras, significa que o ajuste endógeno do modelo está sendo estudado em uma espécie de vácuo econômico no qual não se permite a intrusão de nenhum fator exógeno. É claro que em certos casos esse problema pode não ser muito sério, porque muitos parâmetros econômicos tendem a permane-

cer relativamente constantes por longos períodos de tempo. E em alguns outros casos, podemos realizar um tipo de análise dinâmica comparativa para verificar como a trajetória temporal de uma variável será afetada por uma variação em certos parâmetros. Não obstante, quando estamos interpretando uma trajetória temporal que se estende até um futuro distante, sempre temos de tomar cuidado para não confiar excessivamente na validade da trajetória em seus trechos mais remotos, se tivermos adotado premissas simplificadoras em relação à constância.

Entretanto, é bom que você entenda que o fato de apontarmos essas limitações, como fizemos aqui, não significa, de modo algum, depreciar a análise dinâmica enquanto análise. Na verdade, vale lembrar que, para cada tipo de análise apresentado até aqui, mostramos também suas limitações inerentes. Por conseguinte, contanto que seja devidamente interpretada e adequadamente aplicada, a análise dinâmica – como qualquer outro tipo de análise – pode desempenhar um papel importante no estudo de fenômenos econômicos. Em particular, as técnicas de análise dinâmica nos permitiram a estender, neste capítulo, o estudo da otimização até o reino da otimização dinâmica, na qual a solução que procuramos já não é um estado estacionário ótimo, mas toda uma trajetória temporal ótima.



O alfabeto grego

A	α	alfa
B	β	beta
Γ	γ	gama
Δ	δ	delta
E	ϵ	épsilon
Z	ζ	zeta
H	η	eta
Θ	θ	teta
I	ι	iota
K	κ	capa
Λ	λ	lambda
M	μ	mu, mi
N	ν	nu, ni
Ξ	ξ	csi, xi
O	$\omicron$	ômicron
Π	π	pi
P	ρ	rô
Σ	σ	sigma
T	τ	tau
Υ	υ	ípsilon
Φ	ϕ (ou φ)	fi
X	χ	qui
Ψ	ψ	psi
Ω	ω	ômega

Símbolos matemáticos

1. Conjuntos

$a \in S$	a é um elemento do (pertence ao) conjunto S
$b \notin S$	b não é um elemento do conjunto S
$S \subset T$	o conjunto S é um subconjunto do (está contido no) conjunto T
$T \supset S$	o conjunto T inclui com o conjunto S
$A \cup B$	união do conjunto A com o conjunto B
$A \cap B$	interseção do conjunto A com o conjunto B
$\tilde{S}$	complemento do conjunto S
$\{ \}$ ou $\emptyset$	conjunto vazio
$\{a, b, c\}$	conjunto cujos elementos são a, b e c
$\{x \mid x \text{ tem propriedade } P\}$	conjunto de todos os objetos que têm a propriedade P
$\min\{a, b, c\}$	menor elemento do conjunto especificado

R	conjunto de todos os números reais
R^2	espaço real de duas dimensões
R^n	espaço real de n dimensões
(x, y)	par ordenado
(x, y, z)	tripla ordenada
(a, b)	intervalo aberto de a a b
$[a, b]$	intervalo fechado de a a b

2. Matrizes e Determinantes

A' ou A^T	a transposta da matriz A
A^{-1}	a inversa da matriz A
$ A $	o determinante da matriz A
$ J $	determinante jacobiano
$ H $	determinante hessiano
$ \bar{H} $	determinante hessiano aumentado
$r(A)$	o posto da matriz A
trA	o traço de A
0	matriz nula (matriz zero)
$u \cdot v$	o produto interno (produto escalar) dos vetores u e v
$u'v$	o produto escalar de dois vetores

3. Cálculo

Dada $y = f(x)$, uma função de uma única variável x :

$\lim_{x \rightarrow \infty} f(x)$	o limite de $f(x)$ quando x tende ao infinito
dy	a diferencial primeira de y
d^2y	a diferencial segunda de y
$\frac{dy}{dx}$ ou $f'(x)$	a derivada primeira da função $y = f(x)$
$\left. \frac{dx}{dy} \right _{x=x_0}$ ou $f'(x_0)$	a derivada primeira calculada em $x = x_0$
$\frac{d^2y}{dx^2}$ ou $f''(x)$	a derivada segunda de $y = f(x)$
$\frac{d^ny}{dx^n}$ ou $f^{(n)}(x)$	a derivada n -ésima de $y = f(x)$
$\int f(x) dx$	a integral indefinida de $f(x)$
$\int_a^b f(x) dx$	a integral definida de $f(x)$ de $x = a$ a $x = b$

Dada a função $y = f(x_1, x_2, \dots, x_n)$:

$\frac{\partial y}{\partial x_i}$ ou f_i	a derivada parcial de f em relação a x_i
$\nabla f \equiv \text{grad } f$	o gradiente de f



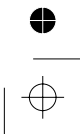
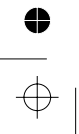
$\frac{dy}{dx_i}$	a derivada total de f em relação a x_i
$\frac{\partial y}{\partial x_i}$	a derivada parcial total de f em relação a x_i

4. Equações diferenciais e de diferenças

$\dot{y} \equiv \frac{dy}{dt}$	a derivada de temporal de y
Δy_t	a primeira diferença de y_t
$\Delta^2 y_t$	a segunda diferença de y_t
y_p	solução particular
y_c	função complementar

5. Outros

$\sum_{i=1}^n x_i$	o somatório de x_i com i variando de 1 a n
$p \Rightarrow q$	p somente se q (p implica q)
$p \Leftarrow q$	p se q (p é implicado por q)
$p \Leftrightarrow q$	p se e somente se q
iff	se e somente se
$ m $	o valor absoluto do número m
$n!$	fatorial de $n \equiv n(n-1)(n-2) \cdots (3)(2)(1)$
$\log_b x$	o logaritmo de x na base b
$\log_e x$ ou $\ln x$	o logaritmo neperiano de x (na base e)
e	a base de logaritmos neperianos e funções exponenciais naturais
$\text{sen } \theta$	função seno de θ
$\text{cos } \theta$	função co-seno de θ
R_n	o termo de resto quando a série de Taylor envolve uma polinômio de n -ésimo grau



Bibliografia recomendada

- ABADIE, J. (Ed.) *Nonlinear Programming*. Amsterdã: North-Holland Publishing Company, 1967. (Uma coletânea de artigos sobre certos aspectos teóricos e computacionais de programação não-linear; o Capítulo 2, de Abadie, trata do teorema de Kuhn-Tucker em relação à qualificação de restrição.)
- ALLEN, R. G. D. *Mathematical Analysis for Economists*. Londres: Macmillan & Co., Ltd., 1938. (Uma exposição clara do cálculo diferencial e integral; são discutidos determinantes, mas não matrizes; nenhuma discussão sobre teoria dos conjuntos, nem sobre programação matemática).
- _____. *Mathematical Economics*. 2. ed. Nova York: St. Martin's Press, Inc., 1959. (Discute uma profusão de modelos matemáticos econômicos; explica equações lineares diferenciais e de diferenças e álgebra matricial.)
- ALMON, C. *Matrix Methods in Economics*. Reading, Mass: Addison-Wesley Publishing Company, Inc., 1967. (São discutidos métodos matriciais em relação a sistemas de equações lineares, modelos insumo-produto, programação linear e programação não-linear. São estudadas também raízes características e vetores característicos.)
- BALDANI, J.; BRADFIELD, J.; TURNER R. *Mathematical Economics*. Orlando: The Dryden Press, 1996.
- BAUMOL, W. J. *Economic Dynamics: An Introduction*. 3. ed. Nova York: The Macmillan Company, 1970. (A Parte IV dá uma explicação lúcida sobre equações de diferenças simples; a Parte V trata de equações de diferenças simultâneas; equações diferenciais são discutidas apenas brevemente.)
- BRAUN, M. *Differential Equations and Their Applications: An Introduction to Applied Mathematics*. 4. ed. Nova York: Springer-Verlag, Inc., 1993. (Contém aplicações interessantes de equações diferenciais, tais como a detecção de falsificação de obras de arte, a disseminação de epidemias, a corrida armamentista e o descarte de resíduos da indústria nuclear.)
- BURMEISTER, E.; DOBELL, A. R. *Mathematical Theories of Economic Growth*. Nova York: The Macmillan Company, 1970. (Uma exposição minuciosa de modelos de crescimento de diversos graus de complexidade.)
- CHIANG, Alpha C. *Elements of Dynamic Optimization*. [S.l.]: McGraw-Hill Book Company, 1992, agora publicado por Waveland Press, Inc., Prospect Heights, Ill.
- CLARK, Colin W. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*. 2. ed. Toronto: John Wiley & Sons, Inc., 1990. (Uma explanação minuciosa da teoria do controle ótimo e sua utilização em recursos renováveis e não-renováveis.)
- CODDINGTON, E. A.; LEVINSON, N. *Theory of Ordinary Differential Equations*. Nova York: McGraw-Hill Book Company, 1955. (Um texto matemático básico sobre equações diferenciais.)
- COURANT, R. *Differential and Integral Calculus* (trans. E. J. McShane). Nova York: Interscience Publishers, Inc., vol. I, 2. ed., 1937, vol. II, 1936. (Um tratado clássico sobre cálculo.)
- _____; John, F. *Introduction to Calculus and Analysis*: Nova York: Interscience Publishers, Inc., vol. I, 1965, vol. II, 1974. (Uma versão atualizada do título anterior.)
- DORFMAN, R.; Samuelson, P. A.; Solow, R. M. *Linear Programming and Economic Analysis*. Nova York: McGraw-Hill Book Company, 1958. (Um tratamento detalhado de programação linear, teoria dos jogos e análise insumo-produção.)
- FRANKLIN, J. *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*. Nova York: Springer-Verlag, Inc., 1980. (Uma deliciosa apresentação da programação matemática.)
- FRISCH, R. *Maxima and Minima: Theory and Economic Applications* (em colaboração com A. Nataf) Chicago, Ill.: Rand McNally & Company, 1966. (Um tratamento completo de problemas de extremos, realizado primariamente segundo a tradição clássica.)
- GOLDBERG, S. *Introduction to Difference Equations*. Nova York: John Wiley & Sons, Inc., 1958. (Com aplicações econômicas.)

- HADLEY, G. *Linear Algebra*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1961. (Aborda matrizes, determinantes, conjuntos convexos etc.)
- . *Linear Programming*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1962. (Uma exposição muito clara, orientada para a matemática.)
- . *Nonlinear and Dynamic Programming*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1964. (Aborda programação não-linear, programação estocástica, programação integral e programação dinâmica; ênfase em aspectos ligados à computação.)
- HALMOS, P. R. *Naive Set Theory*. Princeton, N.J.: D. Van Nostrand Company, Inc., 1960. (Uma introdução informal, portanto fácil de ler, aos aspectos básicos da teoria dos conjuntos).
- HANDS, D. Wade *Introductory Mathematical Economics*. 2. ed. Nova York: Oxford University Press, 2004.
- HENDERSON, J. M.; QUANDT, R. E. *Microeconomic Theory: A Mathematical Approach*. 3. ed. Nova York: McGraw-Hill Book Company, 1980. (Um tratamento matemático abrangente de tópicos microeconômicos.)
- HOY, M.; LIVERNOIS, J.; MCKENNA, C.; REES, R.; STENGOS, T. *Mathematics for Economics*. 2. ed. Cambridge, Mass: The MIT Press, 2001.
- INTRILIGATOR, M. D. *Mathematical Optimization and Economic Theory*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1971. (Uma discussão minuciosa de métodos de otimização, incluindo as técnicas clássicas, programação linear e não-linear e otimização dinâmica; também aplicações às teorias do consumidor e da empresa, equilíbrio geral e economia do bem-estar e teorias de crescimento.)
- KEMENY, J. G.; SNELL, J. L.; THOMPSON, G. L. *Introduction to Finite Mathematics*. 3. ed., Englewood Cliffs, N.J.: Prentice Hall, Inc., 1974. (Abrange tópicos como conjuntos, matrizes, probabilidade e programação linear.)
- KLEIN, Michael W. *Mathematical Methods for Economics*. 2. ed. Reading, Mass: Addison-Wesley Publishing Company, Inc., 2002.
- KOO, D. *Elements of Optimization: With Applications in Economics and Business*. Nova York: Springer-Verlag, Inc., 1977. (Clara discussão de métodos clássicos de otimização, programação matemática, bem como teoria do controle ótimo.)
- KOOPMANS, T. C. (Ed.) *Activity Analysis of Production and Allocation*. Nova York: John Wiley & Sons, Inc., 1951, reimpresso por Yale University Press, 1972. (Contém vários artigos sobre programação linear e análise de atividade.)
- . *Three Essays on the State of Economic Science*. Nova York: McGraw-Hill Book Company, 1957. (O primeiro ensaio contém uma boa exposição de conjuntos convexos; o terceiro ensaio discute a interação de ferramentas e problemas em economia.)
- LAMBERT, Peter J. *Advanced Mathematics for Economists: Static and Dynamic Optimization*. Nova York: Blackwell Publishers, 1985.
- LEONTIEF, W. W. *The Structure of American Economy, 1919–1939*. 2. ed. Fair Lawn, N.J.: Oxford University Press, 1951. (O trabalho pioneiro na análise de insumo-produto.)
- SAMUELSON, P. A. *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press, 1947. (Um clássico da matemática econômica, mas muito difícil de ler.)
- SILBERBERG, Eugene; SUEN, Wing. *The Structure of Economics: A Mathematical Analysis*. 3. ed. Nova York: McGraw-Hill Book Company, 2001. (Primariamente com um foco microeconômico, este livro contém uma forte discussão do teorema do envelope e uma ampla variedade de aplicações.)
- SYDSÆTER, Knut; HAMMOND, Peter. *Essential Mathematics for Economic Analysis*. Londres: Prentice Hall, Inc., 2002.
- TAKAYAMA, A. *Mathematical Economics*. 2. ed. Hinsdale, Ill.: The Dryden Press, 1985. (Dá um tratamento extensivo da teoria econômica em termos matemáticos, concentrando-se em dois tópicos específicos: equilíbrio competitivo e crescimento econômico.)
- THOMAS, G. B.; FINNEY, R. L.: *Calculus and Analytic Geometry*. 9. ed. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1996. (Uma introdução ao cálculo escrita de um modo muito claro.)

Respostas a exercícios selecionados

Exercício 2.3

1. (a) $\{x \mid x > 34\}$
3. (a) $\{2, 4, 6, 7\}$ (c) $\{2, 6\}$ (e) $\{2\}$
8. Há 16 subconjuntos.
9. *Sugestão:* Distinguir entre os dois símbolos $\notin$ e $\nsubseteq$.

Exercício 2.4

1. (a) $\{(3, a), (3, b), (6, a), (6, b), (9, a), (9, b)\}$
3. Não.
5. Imagem = $\{y \mid 8 \leq y \leq 32\}$

Exercício 2.5

2. (a) e (b) são diferentes no sinal do coeficiente angular; (a) e (c) são diferentes na interseção com o eixo vertical.
4. Quando são permitidos valores negativos, o quadrante III também tem de ser usado.
5. (a) x^{19}
6. (a) x^6

Exercício 3.2

1. $P^* = 2\frac{3}{11}$ e $Q^* = 14\frac{2}{11}$
3. *Nota:* Em $2(a)$, $c = 10$ (não 6).
5. *Sugestão:* $b + d = 0$ implica $d = -b$.

Exercício 3.3

1. (a) $x_1^* = 5$ e $x_2^* = 3$
3. (a) $(x-6)(x+1)(x-3) = 0$ ou $x^3 - 8x^2 + 9x + 18 = 0$
5. (a) $-1, 2$ e 3 (c) $-1, \frac{1}{2}$ e $-\frac{1}{4}$

Exercício 3.4

3. $P_1^* = 3\frac{6}{17}$ $P_2^* = 3\frac{8}{17}$ $Q_1^* = 11\frac{7}{17}$ $Q_2^* = 8\frac{7}{17}$

Exercício 3.5

1. (b) $Y^* = (a - bd + I_0 + G_0) / [1 - b(1 - t)]$
 $T^* = [d(1 - b) + t(a + I_0 + G_0)] / [1 - b(1 - t)]$
 $C^* = [a - bd + b(1 - t)(I_0 + G_0)] / [1 - b(1 - t)]$
3. *Sugestão:* Após substituir as últimas duas equações na primeira equação, considere a equação resultante como uma equação quadrática na variável $w \equiv Y^{1/2}$. Somente uma raiz é aceitável, $w_1^* = 11$, dando $Y^* = 121$ e $C^* = 91$. A outra raiz leva a um C^* negativo.

Exercício 4.1

1. Os elementos do vetor (coluna) de constantes são: $0, a, -c$.

Exercício 4.2

1. (a) $\begin{bmatrix} 7 & 3 \\ 9 & 7 \end{bmatrix}$ (c) $\begin{bmatrix} 21 & -3 \\ 18 & 27 \end{bmatrix}$
3. Neste caso especial, por acaso AB é igual a $BA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.
4. (b) $\begin{bmatrix} 49 & 3 \\ 4 & 3 \end{bmatrix}_{(2 \times 2)}$ (c) $\begin{bmatrix} 3x + 5y \\ 4x + 2y - 7z \end{bmatrix}_{(2 \times 1)}$
6. (a) $x_2 + x_3 + x_4 + x_5$ (c) $b(x_1 + x_2 + x_3 + x_4)$
7. (b) $\sum_{i=2}^4 a_i(x_{i+1} + i)$ (d) Sugestão: $x^0 = 1$ se $x \neq 0$

Exercício 4.3

1. (a) $uv' = \begin{bmatrix} 15 & 5 & -5 \\ 3 & 1 & -1 \\ 9 & 3 & -3 \end{bmatrix}$ (c) $xx' = \begin{bmatrix} x_1^2 & x_1 x_2 & x_1 x_3 \\ x_2 x_1 & x_2^2 & x_2 x_3 \\ x_3 x_1 & x_3 x_2 & x_3^2 \end{bmatrix}$
- (e) $u'v = 13$ (g) $u'u = 35$
3. (a) $\sum_{i=1}^n P_i Q_i$ (b) $P \cdot Q$ or $P'Q$ or $Q'P$
5. (a) $2v = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$ (c) $u - v = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$
7. (a) $d = \sqrt{27}$
9. (c) $d(v, 0) = (v \cdot v)^{1/2}$

Exercício 4.4

1. (a) $\begin{bmatrix} 5 & 17 \\ 11 & 17 \end{bmatrix}$
2. Não; deve ser $A - B = -B + A$.
4. (a) $k(A + B) = k[a_{ij} + b_{ij}] = [ka_{ij} + kb_{ij}] = [ka_{ij}] + [kb_{ij}] = k[a_{ij}] + k[b_{ij}] = kA + kB$ (Você consegue justificar cada etapa?)

Exercício 4.5

1. (a) $AI_3 = \begin{bmatrix} -1 & 5 & 7 \\ 0 & -2 & 4 \end{bmatrix}$ (c) $I_2 x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
3. (a) 5×3 (c) 2×1
4. Sugestão: Multiplique a matriz diagonal dada por ela mesma e verifique as condições de idempotência para a matriz produto resultante.

Exercício 4.6

$$1. A' = \begin{bmatrix} 0 & -1 \\ 4 & 3 \end{bmatrix} \text{ e } B' = \begin{bmatrix} 3 & 0 \\ -8 & 1 \end{bmatrix}$$

3. *Sugestão:* Defina $D \equiv AB$, e aplique (4.11).

5. *Sugestão:* Defina $D \equiv AB$, e aplique (4.14).

Exercício 5.1

1. (a) (5,2)

(c) (5,3)

(e) (5,3)

3. (a) Sim.

(d) Não.

5. (a) $r(A) = 3$; A é invertível.

(b) $r(B) = 2$; B é singular.

Exercício 5.2

1. (a) -6

(c) 0

(e) $3abc - a^3 - b^3 - c^3$

$$3. |M_b| = \begin{vmatrix} d & f \\ g & i \end{vmatrix}$$

$$|C_b| = - \begin{vmatrix} d & f \\ g & i \end{vmatrix}$$

4. (a) *Sugestão:* Expanda pela terceira coluna.

5. 20 (não -20)

Exercício 5.3

3. (a) Propriedade IV.

(b) Propriedade III (aplicada a ambas as linhas).

4. (a) Singular.

(c) Singular.

5. (a) Posto < 3

(c) Posto < 3

7. A é invertível porque $|A| = 1 - b \neq 0$.

Exercício 5.4

$$1. \sum_{i=1}^4 a_{i3} |C_{22}| \quad \sum_{j=1}^4 a_{2j} |C_{4j}|$$

3. (a) Permute os dois elementos da diagonal de A ; multiplique os dois elementos de fora da diagonal de A por -1.

(b) Divida por $|A|$.

$$4. (a) E^{-1} = \frac{1}{20} \begin{bmatrix} 3 & 2 & -3 \\ -7 & 2 & 7 \\ -6 & -4 & 26 \end{bmatrix} \quad (c) G^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Exercício 5.5

1. (a) $x_1^* = 4$ e $x_2^* = 3$

(c) $x_1^* = 2$ e $x_2^* = 1$

$$2. (a) A^{-1} = \frac{1}{7} \begin{bmatrix} 1 & 2 \\ -2 & 3 \end{bmatrix}; x^* = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad (c) A^{-1} = \frac{1}{15} \begin{bmatrix} 1 & 7 \\ -1 & 8 \end{bmatrix}; x^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

3. (a) $x_1^* = 2$, $x_2^* = 0$, $x_3^* = 1$

(c) $x^* = 0$, $y^* = 3$, $z^* = 4$

4. *Sugestão:* Aplique (5.8) e (5.13).

Exercício 5.6

$$1. (a) A^{-1} = \frac{1}{1-b+bt} \begin{bmatrix} 1 & 1 & -b \\ b(1-t) & 1 & -b \\ t & t & 1-b \end{bmatrix}$$

$$\begin{bmatrix} Y^* \\ C^* \\ T^* \end{bmatrix} = \frac{1}{1-b+bt} \begin{bmatrix} I_0 + G_0 + a - bd \\ b(1-t)(I_0 + G_0) + a - bd \\ t(I_0 + G_0) + at + d(1-b) \end{bmatrix}$$

$$(b) |A| = 1 - b + bt$$

$$|A_1| = I_0 + G_0 - bd + a$$

$$|A_2| = a - bd + b(1-t)(I_0 + G_0)$$

$$|A_3| = d(1-b) + t(a + I_0 + G_0)$$

Exercício 5.7

$$1. x_1^* = 69,53, x_2^* = 57,03 \text{ e } x_3^* = 42,58$$

$$3. (a) A = \begin{bmatrix} 0,10 & 0,50 \\ 0,60 & 0 \end{bmatrix}; \text{ a equação matricial é } \begin{bmatrix} 0,90 & -0,50 \\ -0,60 & 1,00 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1000 \\ 2000 \end{bmatrix}.$$

$$(c) x_1^* = 3.333\frac{1}{3} \text{ e } x_2^* = 4.000$$

4. Elemento 0,33: são necessários 33¢ da Mercadoria II como insumo para produzir \$1 da Mercadoria I.

Exercício 6.2

$$1. (a) \Delta y / \Delta x = 8x + 4\Delta x \quad (b) dy/dx = 8x \quad (c) f'(3) = 24, f'(4) = 32$$

$$3. (a) \Delta y / \Delta x = 5; \text{ uma função constante.}$$

Exercício 6.4

$$1. \text{ Limite do lado esquerdo = limite do lado direito = 15; o limite é 15.}$$

$$3. (a) 5 \quad (b) 5$$

Exercício 6.5

$$1. (a) -3/4 < x \quad (c) x < 1/2$$

$$3. (a) -7 < x < 5 \quad (c) -4 \leq x \leq 1$$

Exercício 6.6

$$1. (a) 7 \quad (c) 17$$

$$3. (a) 2\frac{1}{2} \quad (c) 2$$

Exercício 6.7

$$2. (a) N^2 - 5N - 2 \quad (b) \text{ Sim.}$$

$$(c) \text{ Sim.}$$

$$3. (a) (N+2)/(N^2+2) \quad (b) \text{ Sim.}$$

$$(c) \text{ Contínua no domínio.}$$

$$6. \text{ Sim; cada função é contínua e suave.}$$

Exercício 7.1

1. (a) $dy/dx = 12x^{11}$ (c) $dy/dx = 35x^4$ (e) $dw/du = -2u^{-1/2}$
3. (a) $f'(x) = 18$; $f'(1) = f'(2) = 18$
 (c) $f'(x) = 10x^{-3}$; $f'(1) = 10$, $f'(2) = 1\frac{1}{4}$

Exercício 7.2

1. $VC = Q^3 - 5Q^2 + 12Q$; $\frac{dVC}{dQ} = 3Q^2 - 10Q + 12$ é a função MC.
3. (a) $3(27x^2 + 6x - 2)$ (c) $12x(x + 1)$ (e) $-x(9x + 14)$
4. (b) $MR = 60 - 6Q$
7. (a) $(x^2 - 3)/x^2$ (c) $30/(x + 5)^2$
8. (a) a (c) $-a/(ax + b)^2$

Exercício 7.3

1. $-2x[3(5 - x^2)^2 + 2]$
3. (a) $18x(3x^2 - 13)^2$ (c) $5a(ax + b)^4$
5. $x = \frac{1}{7}y - 3$, $dx/dy = \frac{1}{7}$

Exercício 7.4

1. (a) $\partial y/\partial x_1 = 6x_1^2 - 22x_1x_2$ e $\partial y/\partial x_2 = -11x_1^2 + 6x_2$
 (c) $\partial y/\partial x_1 = 2(x_2 - 2)$ e $\partial y/\partial x_2 = 2x_1 + 3$
3. (a) 12 (c) 10/9
5. (a) $U_1 = 2(x_1 + 2)(x_2 + 3)^3$ e $U_2 = 3(x_1 + 2)^2(x_2 + 3)^2$

Exercício 7.5

1. $\partial Q^*/\partial a = d/(b + d) > 0$ $\partial Q^*/\partial b = -d(a + c)/(b + d)^2 < 0$
 $\partial Q^*/\partial c = -b/(b + d) < 0$ $\partial Q^*/\partial d = b(a + c)/(b + d)^2 > 0$
2. $\partial Y^*/\partial I_0 = \partial Y^*/\partial \alpha = 1/(1 - \beta + \beta\delta) > 0$

Exercício 7.6

1. (a) $|J| = 0$; as funções são dependentes.
 (b) $|J| = -20x_2$; as funções são independentes.

Exercício 8.1

1. (a) $dy = -3(x^2 + 1) dx$ (c) $dy = [(1 - x^2)/(x^2 + 1)^2] dx$
3. (a) $dC/dY = b$ e $C/Y = (a + bY)/Y$

Exercício 8.2

2. (a) $dz = (6x + y) dx + (x - 6y^2) dy$
3. (a) $dy = [x_2/(x_1 + x_2)^2] dx_1 - [x_1/(x_1 + x_2)^2] dx_2$
4. $\varepsilon_{QP} = 2bP^2/(a + bP^2 + R^{1/2})$
6. $\varepsilon_{XP} = -2/(Y_f^{1/2} P^2 + 1)$

Exercício 8.3

3. (a) $dy = 3[(2x_2 - 1)(x_3 + 5)dx_1 + 2x_1(x_3 + 5)dx_2 + x_1(2x_2 - 1)dx_3]$
4. *Sugestão:* Aplique as definições de diferencial e diferencial total.

Exercício 8.4

1. (a) $dz/dy = x + 10y + 6y^2 = 28y + 9y^2$
(c) $dz/dy = -15x + 3y = 108y - 30$
3. $dQ/dt = [a\alpha A/K + b\beta A/L + A'(t)]K^\alpha L^\beta$
4. (b) $\$W/\$u = 10uf_1 + f_2 \quad \$W/\$v = 3f_1 - 12v^2f_2$

Exercício 8.5

5. (a) Definida; $dy/dx = -(3x^2 - 4xy + 3y^2)/(-2x^2 + 6xy) = -9/8$
(b) Definida; $dy/dx = -(4x + 4y)/(4x - 4y^3) = 2/13$
7. A condição $F_y \neq 0$ é violada em $(0, 0)$.
8. O produto das derivadas parciais é igual a -1 .

Exercício 8.6

1. (c) $(dY^*/dG_0) = 1/(S' + T' - I') > 0$
3. $(\partial P^*/\partial Y_0) = D_{Y_0}/(S_{P^*} - D_{P^*}) > 0 \quad (\partial Q^*/\partial Y_0) = D_{Y_0} S_{P^*}/(S_{P^*} - D_{P^*}) > 0$
 $(\partial P^*/\partial T_0) = -S_{T_0}/(S_{P^*} - D_{P^*}) > 0 \quad (\partial Q^*/\partial T_0) = -S_{T_0} D_{P^*}/(S_{P^*} - D_{P^*}) < 0$

Exercício 9.2

1. (a) Quando $x = 2, y = 15$ (um máximo relativo).
(c) Quando $x = 0, y = 3$ (um mínimo relativo).
2. (a) O valor crítico $x = -1$ está fora do domínio; o valor crítico $x = 1$ leva a $y = 3$ (um mínimo relativo).
4. (d) A elasticidade é um.

Exercício 9.3

1. (a) $f''(x) = 2a, f'''(x) = 0 \quad (c) f''(x) = 6(1 - x)^{-3}, f'''(x) = 18(1 - x)^{-4}$
3. (b) Uma linha reta.
5. Cada um dos pontos sobre $f(x)$ é um ponto estacionário, mas o único ponto estacionário sobre $g(x)$ que conhecemos é em $x = 3$.

Exercício 9.4

1. (a) $f(2) = 33$ é um máximo.
(c) $f(1) = 5\frac{1}{3}$ é um máximo; $f(5) = -5\frac{1}{3}$ é um mínimo.
2. *Sugestão:* Primeiro escreva uma função área A em termos de uma só variável (seja L ou W).
3. (d) $Q^* = 11 \quad (e) \text{ Lucro máximo} = 111\frac{1}{3}$
5. (a) $k < 0 \quad (b) h < 0 \quad (c) j > 0$
7. (b) S é maximizada no nível de produção 20,37 (aproximadamente).

Exercício 9.5

1. (a) 120 $(c) 4 \quad (e) (n + 2)(n + 1)$
2. (a) $1 + x + x^2 + x^3 + x^4$
3. (b) $-63 - 98x - 62x^2 - 18x^3 - 2x^4 + R_4$

Exercício 9.6

1. (a) $f(0) = 0$ é um ponto de inflexão. (c) $f(0) = 5$ é um mínimo relativo.
2. (b) $f(2) = 0$ é um mínimo relativo.

Exercício 10.1

1. (a) Sim. (b) Sim.
3. (a) $5e^{5t}$ (c) $-12e^{-2t}$
5. (a) A curva com $a = -1$ é a imagem especular da curva com $a = 1$ em relação ao eixo horizontal.

Exercício 10.2

1. (a) 7,388 (b) 1,649
2. (c) $1 + 2x + \frac{1}{2!}(2x)^2 + \frac{1}{3!}(2x)^3 + \dots$
3. (a) $\$70e^{0,12}$ (b) $\$690e^{0,10}$

Exercício 10.3

1. (a) 4 (c) 4
2. (a) 7 (c) -3 (e) 6
3. (a) 26 (c) $\ln 3 - \ln B$ (f) 3

Exercício 10.4

1. O requisito impede que a função degenera para uma função constante.
3. *Sugestão:* Tome o logaritmo na base b .
4. (a) $y = e^{(3 \ln 8)t}$ ou $y = e^{6,2385t}$ (c) $y = 5e^{(\ln 5)t}$ ou $y = 5e^{1,6095t}$
5. (a) $t = (\ln y)/(\ln 7)$ ou $t = 0,5139 \ln y$
(c) $t = 3 \ln(9y)/(\ln 15)$ ou $t = 1,1078 \ln(9y)$
6. (a) $r = \ln 1,05$ (c) $r = 2 \ln 1,03$

Exercício 10.5

1. (a) $2e^{2t+4}$ (c) $2te^{t^2+1}$ (e) $(2ax + b)e^{ax^2+bx+c}$
3. (a) $5/t$ (c) $1/(t+19)$ (e) $1/[x(1+x)]$
5. *Sugestão:* Use (10.21) e aplique a regra da cadeia.
7. (a) $3(8-x^2)/[(x+2)^2(x+4)^2]$

Exercício 10.6

1. $t^* = 1/r^2$
2. $d^2A/dt^2 = -A(\ln 2)/4\sqrt{t^3} < 0$

Exercício 10.7

1. (a) $2/t$ (c) $\ln b$ (e) $1/t - \ln 3$
3. $r_y = kr_x$
7. $|\varepsilon_d| = n$
11. $r_Q = \varepsilon_{QK}r_K + \varepsilon_{QL}r_L$

Exercício 11.2

1. $z^* = 3$ é um mínimo.
3. $z^* = c$, que é um mínimo no caso (a), um máximo no caso (b) e um ponto de sela no caso (c).
5. (a) Qualquer par (x, y) exceto $(2, 3)$ resulta em um valor positivo de z .
(b) Sim. (c) Não. (d) Sim ($d^2z = 0$).

Exercício 11.3

1. (a) $q = 4u^2 + 4uv + 3v^2$ (c) $q = 5x^2 + 6xy$
3. (a) Positiva definitiva. (c) Nenhuma das duas.
5. (a) Positiva definitiva. (c) Positiva definitiva. (e) Positiva definitiva.
6. (a) $r_1, r_2 = \frac{1}{2}(7 \pm \sqrt{17})$; $u'Du$ é positiva definitiva.
(c) $r_1, r_2 = \frac{1}{2}(5 \pm \sqrt{61})$; $u'Fu$ é indefinida.

$$7. v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Exercício 11.4

1. $z^* = 0$ (mínimo)
3. $z^* = -11/40$ (mínimo)
4. $z^* = 2 - e$ (mínimo), atingido em $(x^*, y^*, w^*) = (0, 0, 1)$
5. (b) *Sugestão*: Veja (11.16).
6. (a) $r_1 = 2$ $r_2 = 4 + \sqrt{6}$ $r_3 = 4 - \sqrt{6}$

Exercício 11.5

1. (a) Estritamente convexa. (c) Estritamente convexa.
2. (a) Estritamente côncava. (c) Nenhuma.
3. Não.
5. (a) Disco. (b) Sim.
7. (a) Combinação convexa, com $\theta = 0,5$. (b) Combinação convexa, com $\theta = 0,2$.

Exercício 11.6

1. (a) Não. (b) $Q_1^* = P_{10}/4$ e $Q_2^* = P_{20}/4$
3. $|\varepsilon_{d1}| = 1\frac{5}{8}$ $|\varepsilon_{d2}| = 1\frac{1}{3}$ $|\varepsilon_{d3}| = 1\frac{1}{2}$
5. (a) $\pi = P_0 Q(a, b) (1 + \frac{1}{2} i_0)^{-2} - P_{a0}a - P_{b0}b$

Exercício 11.7

1. $(\partial a^*/\partial P_{a0}) = P_0 Q_{bb} e^{-rt} / J < 0$ $(\partial b^*/\partial P_{a0}) = -P_0 Q_{ab} e^{-rt} / J < 0$
2. (a) Quatro (b) $(\partial a^*/\partial P_0) = (Q_b Q_{ab} - Q_a Q_{bb}) P_0 (1 + i_0)^{-2} / J > 0$
(c) $(\partial a^*/\partial i_0) = (Q_a Q_{bb} - Q_b Q_{ab}) P_0^2 (1 + i_0)^{-3} / J < 0$

Exercício 12.2

1. (a) $z^* = 1/2$, atingido quando $\lambda^* = 1/2$, $x^* = 1$ e $y^* = 1/2$
(c) $z^* = -19$, atingido quando $\lambda^* = -4$, $x^* = 1$ e $y^* = 5$
4. $Z_{\lambda_i} = -G(x, y) = 0$ $Z_x = f_x - \lambda G_x = 0$ $Z_y = f_y - \lambda G_y = 0$
5. *Sugestão*: Faça a distinção entre igualdade idêntica e igualdade condicional.

Exercício 12.3

1. (a) $|\bar{H}| = 4$; z^* é um máximo. (c) $|\bar{H}| = -2$; z^* é um mínimo.

Exercício 12.4

2. (a) Quase-côncava, mas não estritamente. (c) Estritamente quase-côncava.
 4. (a) Nenhuma. (c) Quase-convexa, mas não quase-côncava.
 5. *Sugestão:* Volte à Seção 9.4.
 7. *Sugestão:* Use ou (12.21) ou (12.25').

Exercício 12.5

1. (b) $\lambda^* = 3$, $x^* = 16$, $y^* = 11$ (c) $|\bar{H}| = 48$; a condição é satisfeita.
 3. $(\partial x^*/\partial B) = 1/2P_x > 0$ $(\partial x^*/\partial P_x) = -(B + P_y)/2P_x^2 < 0$
 $(\partial x^*/\partial P_y) = 1/2P_x > 0$ etc.
 5. Não é válida.
 7. Não para ambos (a) e (b) – veja (12.32) e (12.33').

Exercício 12.6

1. (a) Homogênea de grau um. (c) Não-homogênea.
 (e) Homogênea de grau dois.
 4. São verdadeiras.
 7. (a) Homogênea de grau $a + b + c$.
 8. (a) $f^2 Q = g(jK, jL)$ (b) *Sugestão:* Faça $j = 1/L$.
 (d) Homogênea de grau um em K e L .

Exercício 12.7

1. (a) $1 : 2 : 3$ (b) $1 : 4 : 9$
 2. *Sugestão:* Volte às Figuras 8.2 e 8.3.
 4. *Sugestão:* Esta é uma derivada total.
 6. (a) Retas com inclinações descendentes. (b) $\sigma \rightarrow \infty$ quando $\rho \rightarrow -1$
 8. (a) 7 (c) $\ln 5 - 1$

Exercício 13.1

3. As condições $x_j(\partial Z/\partial x_j) = 0$ e as condições $\lambda_i(\partial Z/\partial \lambda_i) = 0$ podem ser condensadas.
 5. Consistente.

Exercício 13.2

1. Nenhum arco qualificador pode ser encontrado para um vetor teste como $(dx_1, dx_2) = (1, 0)$.
 3. $(x_1^*, x_2^*) = (0, 0)$ é uma cúspide. A qualificação de restrição é satisfeita (todos os vetores teste são horizontais e apontam na direção leste); as condições de Kuhn-Tucker também são satisfeitas.
 4. Todas as condições podem ser satisfeitas escolhendo $y_0^* = 0$ e $y_1^* \geq 0$.

Exercício 13.4

2. (a) Sim. (b) Sim. (c) Não.
 4. (a) Sim. (b) Sim.

Exercício 14.2

1. (a) $-8x^{-2} + c, (x \neq 0)$ (c) $\frac{1}{6}x^6 - \frac{3}{2}x^2 + c$
2. (a) $13e^x + c$ (c) $5e^x - 3x^{-1} + c, (x \neq 0)$
3. (a) $3 \ln|x| + c, (x \neq 0)$ (c) $\ln(x^2 + 3) + c$
4. (a) $\frac{2}{3}(x+1)^{3/2}(x+3) - \frac{4}{15}(x+1)^{5/2} + c$

Exercício 14.3

1. (a) $4\frac{1}{3}$ (b) $3\frac{1}{4}$ (e) $2\left(\frac{a}{3} + c\right)$
2. (a) $\frac{1}{2}(e^{-2} - e^{-4})$ (c) $e^2\left(\frac{1}{2}e^4 - \frac{1}{2}e^2 + e - 1\right)$
3. (b) Subestimativa. (e) $f(x)$ é integrável por Riemann.

Exercício 14.4

1. Nenhuma.
2. (a), (c), (d) e (e).
3. (a), (c) e (d) convergentes; (e) divergente.

Exercício 14.5

1. (a) $R(Q) = 14Q^2 - \frac{10}{3}e^{0,3Q} + \frac{10}{3}$ (b) $R(Q) = 10Q/(1+Q)$
3. (a) $K(t) = 9t^{4/3} + 25$
5. (a) 29.000

Exercício 14.6

1. Somente o capital é considerado. Uma vez que o trabalho normalmente também é necessário para a produção, a premissa subjacente é que K e L são sempre usados segundo uma proporção fixa.
3. *Sugestão:* Use (6.8).
4. *Sugestão:* $\ln u - \ln v = \ln \frac{u}{v}$

Exercício 15.1

1. (a) $y(t) = -e^{-4t} + 3$ (c) $y(t) = \frac{3}{2}(1 - e^{-10t})$
3. (a) $y(t) = 4(1 - e^{-t})$ (c) $y(t) = 6e^{5t}$ (e) $y(t) = 8e^{7t} - 1$

Exercício 15.2

1. A curva D deve ser mais inclinada.
3. O mecanismo de ajuste de preço gera uma equação diferencial.
5. (a) $P(t) = A \exp\left(\frac{\beta + \delta}{\eta}\right) + \frac{\alpha}{\beta + \delta}$ (b) Sim.

Exercício 15.3

1. $y(t) = Ae^{-5t} + 3$
3. $y(t) = e^{-t^2} + \frac{1}{2}$
5. $y(t) = e^{-6t} - \frac{1}{7}e^t$
6. *Sugestão:* Volte à Seção 14.2, Exemplo 17.

Exercício 15.4

1. (a) $y(t) = (c/t^3)^{1/2}$ (c) $yt + y^2t = c$

Exercício 15.5

1. (a) Separável; linear quando escrita como $\frac{dy}{dt} + \frac{1}{t}y = 0$
 (c) Separável; redutível a uma equação de Bernoulli.
 3. $y(t) = (A - t^2)^{1/2}$

Exercício 15.6

1. (a) Linha de fase com inclinação ascendente; equilíbrio dinamicamente instável.
 (c) Linha de fase com inclinação descendente; equilíbrio dinamicamente estável.
 3. O sinal da derivada mede a inclinação da linha de fase.

Exercício 15.7

1. $r_k = r_K - r_L$ (10.25)
 4. (a) Faça os gráficos de $(3 - y)$ e $\ln y$ como duas curvas separadas e depois subtraia. Existe um único equilíbrio (a um valor de y entre 1 e 3) ele é dinamicamente estável.

Exercício 16.1

1. (a) $y_p = 2/5$ (c) $y_p = 3$ (e) $y_p = 6t^2$
 3. (a) $y(t) = 6e^t + e^{-4t} - 3$ (c) $y(t) = e^t + te^t + 3$
 6. *Sugestão:* Aplique a regra de L'Hôpital.

Exercício 16.2

1. (a) $\frac{3}{2} \pm \frac{3}{2}\sqrt{3}i$ (c) $-\frac{1}{4} \pm \frac{3}{4}\sqrt{7}i$
 3. (b) *Sugestão:* Quando $\theta = \pi/4$, a reta OP é uma reta a 45° .
 5. (a) $\frac{d}{d\theta} \sin f(\theta) = f'(\theta) \cos f(\theta)$ (b) $\frac{d}{d\theta} \cos \theta^3 = -3\theta^2 \sin \theta^3$
 7. (a) $\sqrt{3} + i$ (c) $1 - i$

Exercício 16.3

1. $y(t) = e^{2t}(3 \cos 2t + \frac{1}{2} \sin 2t)$
 3. $y(t) = e^{-3t/2} \left(-\cos \frac{\sqrt{7}}{2}t + \frac{\sqrt{7}}{7} \sin \frac{\sqrt{7}}{2}t \right) + 3$
 5. $y(t) = \frac{2}{3} \cos 3t + \sin 3t + \frac{1}{3}$

Exercício 16.4

1. (a) $P'' + \frac{m-u}{n-w}P' - \frac{\beta+\delta}{n-w} = -\frac{\alpha+\gamma}{n-w}$ ($n \neq w$) (b) $P_p = \frac{\beta+\gamma}{\beta+\delta}$
 3. (a) $P(t) = e^{t/2} \left(2 \cos \frac{3}{2}t + 2 \sin \frac{3}{2}t \right) + 2$

Exercício 16.5

1. (a) $\frac{d\pi}{dt} + j(1 - g) = j(\alpha - T - \beta U)$
 (b) Nenhuma raiz complexa; nenhuma flutuação.
3. (c) Ambas são equações diferenciais de primeira ordem. (d) $g \neq 1$
4. (a) $\pi(t) = e^{-t} \left(A_5 \cos \frac{\sqrt{2}}{4} t + A_6 \sin \frac{\sqrt{2}}{4} t \right) + m$ (c) $\bar{P} = m; \bar{U} = \frac{1}{18} - \frac{2}{9} m$

Exercício 16.6

2. (a) $y_p = t - 2$ (c) $y_p = \frac{1}{4} e^t$

Exercício 16.7

1. (a) $y_p = 4$ (c) $y_p = \frac{1}{18} t^2$
3. (a) Divergente. (c) Convergente.

Exercício 17.2

1. (a) $y_{t+1} = y_t + 7$ (c) $y_{t+1} = 3y_t - 9$
3. (a) $y_t = 10 + t$ (c) $y_t = y_0 \alpha^t - \beta(1 + \alpha + \alpha^2 + \dots + \alpha^{t-1})$

Exercício 17.3

1. (a) Não-oscilatória; divergente. (c) Oscilatória; convergente.
3. (a) $y_t = -8(1/3)^t + 9$ (c) $y_t = -2(-1/4)^t + 4$

Exercício 17.4

1. $Q_t = \alpha - \beta(P_0 - \bar{P})(-\delta/\beta)^t - \beta\bar{P}$
3. (a) $\bar{P} = 3$; oscilação explosiva. (c) $\bar{P} = 2$; oscilação uniforme.
5. A defasagem na função oferta.

Exercício 17.5

1. $a = -1$
3. $P_t = (P_0 - 3)(-1,4)^t + 3$, com oscilação explosiva.

Exercício 17.6

1. Não.
2. (b) Movimento descendente explosivo, não-oscilatório.
 (d) Movimento descendente constante, amortecido em direção a R .
4. (a) No início, inclinação descendente, em seguida, torna-se horizontal.

Exercício 18.1

1. (a) $\frac{1}{2} \pm \frac{1}{2} i$ (c) $\frac{1}{2}, -1$
3. (a) 4 (constante) (c) 5 (constante)
4. (b) $y_t = \sqrt{2}^t \left(2 \cos \frac{\pi}{4} t + \sin \frac{\pi}{4} t \right) + 1$

Exercício 18.2

1. (a) Subcaso 1D. (c) Subcaso 1C.
3. *Sugestão:* Use (18.16).

Exercício 18.3

3. Possibilidades v , $viii$, x e xi se tornarão viáveis.
4. (a) $p_{t+2} - [2 - j(1 - g) - \beta k]p_{t+1} + [1 - j(1 - g) - \beta k(1 - j)]p_t = j\beta km$
 (c) $\beta k \gtrless 4$

Exercício 18.4

1. (a) 1 (c) $3t^2 + 3t + 1$
3. (a) $y_p = \frac{1}{4}t$ (c) $y_p = 2 - t + t^2$
5. (a) $1/2, -1$ e 1

Exercício 19.2

2. $b^3 + b^2 - 3b + 2 = 0$
3. (a) $x_t = -(3)^t + 4(-2)^t + 7$ $y_t = 2(3)^t + 2(-2)^t + 5$
4. (a) $x(t) = 4e^{-2t} - 3e^{-3t} + 12$ $y(t) = -e^{-2t} + e^{-3t} + 4$

Exercício 19.3

2. (c) $\beta = (\delta I - A)^{-1}u$
3. (c) $\beta = (\rho I + I - A)^{-1}\lambda$
5. (c) $x_1(t) = 4e^{-4t/10} + 2e^{-11t/10} + \frac{17}{6}e^{t/10}$; $x_2(t) = 3e^{-4t/10} - 2e^{-11t/10} + \frac{19}{6}e^{t/10}$

Exercício 19.4

4. (a)
$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = \left[\frac{A_1}{23 - \sqrt{193}} A_1 \right] \left(\frac{33 + \sqrt{193}}{64} \right)^t + \left[\frac{A_2}{23 + \sqrt{193}} A_2 \right] \left(\frac{33 - \sqrt{193}}{64} \right)^t + \begin{bmatrix} \mu \\ \frac{1}{6}(1 - \mu) \end{bmatrix}$$

Exercício 19.5

1. A equação única pode ser reescrita como duas equações de primeira ordem.
2. Sim.
4. (a) Ponto de sela.

Exercício 19.6

1. (a) $|J_E| = 1$ e $\text{tr } J_E = 2$; nó localmente instável.
 (c) $|J_E| = 5$ e $\text{tr } J_E = -1$; foco localmente estável.
2. (a) Localmente um ponto de sela. (c) Nó localmente estável ou foco estável.
4. (a) As curvas $x' = 0$ e $y' = 0$ coincidem e resultam em uma reta cheia de pontos de equilíbrio.

Exercício 20.2

1. $\lambda^* = 1 - t$

$$u^* = \frac{1-t}{2} \quad y^* = \frac{t}{2} - \frac{t^2}{4} + 2$$

6. $\lambda^*(t) = 3e^{4-t} - 3$

$$u^*(t) = 2 \quad y^*(t) = 7e^t - 2$$

Exercício 20.4

1. $\lambda^* = \delta / (\delta^2 + \alpha)$

$$K^* = 1/2(\delta^2 + \alpha)$$

Índice

A

Abordagem de equações simultâneas, 197-199
Abscissa, 37
Acelerador, interação com multiplicador, 552-557
Acordo oficial, mudança de, 204n
Adjunta, 98
Alfabeto grego, 631
Amplitude, 495
Análise de equilíbrio. *Veja* Análise estática
Análise de equilíbrio geral, 43
Análise de período, 523
Análise dinâmica, limitações da, 628
Análise estática
 limitações da, 118
 modelos de insumo-produção de Leontief, 110-118
Antilog, 451
Aposta justa, 221
Aproximação linear de uma função, 234-236
Arco qualificador, 394, 395
Área negativa, 438
Área sob uma curva, 436-438
Argumento, 19
Arranjos, matrizes como, 50-51
Arrow, K. J., 350n, 376n, 404n
Assíntota, 22
Autovalor, 292n
Autovetor, 292n

B

Balança de pagamentos, 204
Base, 63
 de função exponencial, 244, 246
 de função logarítmica, 254-256
Bem inferior, 359
Bem normal, 359
Bens de Giffen, 361
Boltyanskii, V. G., 609n

C

Cadeias de Markov, 77-80
 absorventes, 80
 finitas, 79
Cálculo de variações, 607
Cálculo diferencial, 122
Cálculo integral, 426

Caminho de fase, 593
Caminho, 593
Capital
 dinâmica do, 477-481
 investimento e, 445-447
cartesiana, 498, 549
 polar, 499
Casos de múltiplas restrições, 336, 344
Chenery, H. B., 376n
Chiang, A. C., 4, 287n, 607n
Círculo unitário, 502
Coeficiente(s), 8
 constante, 483
 de aceleração, 552-553
 de ajuste, 460
 de insumo, 110
 de utilização, 452
 fracionário, 39
 indeterminado, 516-518, 562-565, 581, 583
Coeficientes constantes, 483
Co-fator
 espúrio, 97
 definição, 89
Combinação convexa, 311-313
Combinação de insumos de custo mínimo, 370-380
Combinação linear, 61, 62
Complemento de conjunto, Conjunto complemento, 13
Completar o quadrado, 37-38, 228n, 288, 290
Composição de juro, 250
Compressão, 245, 260
Conceito de estoque, 251, 446
Conceito de fluxo, 251, 446
Conceito pontual de tempo, 251
Condição de equilíbrio, 9
Condição de fronteira, 426
Condição de Hawkins-Simon, 114
 menor principal e, 289, 290, 291, 298
 significado econômico da, 116
Condição de positiva definitiva e negativa definitiva, 287, 291, 292, 296
Condição de primeira ordem, 223, 280, 381
 derivada *versus* forma diferencial de, 277-278, 279
 necessária *versus* suficiente, 280
 para extremo, 297

Condição de segunda ordem, 283-285, 297-301
 em relação à concavidade e à convexidade, 302-314
 em relação à quase-concavidade e à quase-convexidade, 345-355
 forma derivada *versus* forma diferencial de, 278-279
 necessária *versus* suficiente, 224, 283, 284, 339
 papel da, em estática comparativa, 327
Condição de transversalidade, 610, 612, 615
Condição inicial, 426
Condição necessária, 81-82, 224, 226, 339, 403
Condição necessária e suficiente, 82, 83, 403
Condição subsidiária, 330. *Veja também* Restrição
Condição suficiente, 81-82, 224, 339, 403
Condições de derivadas *versus* condições de diferenciais, 277-279
Condições de Kuhn-Tucker, 381-391
 efeitos das restrições de desigualdade, 383-387
 interpretação econômica de, 387-388
 teoria do controle ótimo e, 615
 versão de minimização de, 388
Condições de otimização, 9
Condições de reciprocidade, 408-410
Condições terminais alternativas, 615-619
Conjunto convexo *versus* função convexa, 310-313
Conjunto(s), 10-15
 complemento de, 13
 disjunto, 12-13
 enumerável *versus* finito *versus* infinito, 11
 igualdade de, 12
 interseção de, 12-13
 leis de operação com, 14-16
 não-enumerável, 11
 nulo, 12
 operações com, 12-15
 ordenado, 16
 relações entre, 11-13
 subconjunto, 12

- união de, 12-13
 - universal, 13
 - vazio, 12
 - Constante(s), 287
 - aditiva, 148
 - de integração, 427
 - definição, 8
 - expoentes como, 244
 - multiplicativa, 148
 - paramétrica, 8
 - Contar equações e incógnitas, 44
 - Continuidade, 137
 - de função derivada, 149
 - de função polinomial, 138
 - de função racional, 138
 - em relação à diferenciabilidade, 138-142
 - Controle ótimo
 - ilustração de, 608-609
 - natureza de, 607-614
 - Convergência, 542
 - de integral imprópria, 441-444
 - de séries, 237, 248
 - divergência *versus*, 554-557
 - Conversão de base, 261-262
 - Coordenadas cartesianas, 498, 549
 - Coordenadas polares, 499
 - Correspondência um para um, 17, 60, 157, 159
 - Courant, R., 241n
 - Crescimento
 - contínuo *versus* discreto, 252-253
 - funções exponenciais e, 248-254
 - lei exponencial de, 243
 - modelo de Domar de, 450-453, 455
 - modelo de Solow de, 477-481, 626
 - modelo neoclássico ótimo de, 624-626
 - negativo, 253
 - taxa de, 251-252, 272-274
 - taxa instantânea de, 251-252, 272-274
 - CRTS. *Veja* Retornos constantes de escala (Constant returns to scale - CRTS)
 - Curva de indiferença, 355-358
 - Curvas de demarcação, 591-593
 - Curvas de isovalor, 372n
 - Cúspide, 392, 393
 - Custo(s)
 - marginal *versus* total, 124-125, 148
 - médio *versus* marginal, 154
 - minimização de, 370-380
 - Custo marginal
 - custo médio *versus*, 154
 - custo total *versus*, 124-125, 148, 444-445
 - Custo médio *versus* custo marginal, 154
- D**
- Decisão de insumos, 319-324
 - Defasagem
 - em consumo, 553
 - em oferta, 533
 - em produção, 580-582
 - Definição de sinal
 - positivo e negativo, 287
 - teste com determinantes para, 287-289
 - teste da raiz característica para, 292-296
 - Demanda, 32, 33, 36
 - com expectativas de preço, 506
 - de insumo, 111
 - elasticidade de, 179, 318-319
 - excesso de. *Veja* Excesso de demanda
 - final, 111
 - funções demanda hicksianas, 415
 - marshalliana, 414, 415-418
 - receita média e, 315-316
 - Dependência
 - entre colunas e linhas de uma matriz, 94
 - entre equações, 44-45, 84
 - linear, 62-63
 - Depreciação, taxa de, 253
 - Derivação, 139
 - Derivada(s), 123
 - continuidade de, 149
 - da função co-seno, 496
 - de estática comparativa. *Veja* Estática comparativa
 - de funções exponenciais, 264-266
 - derivada de, 217-219
 - função marginal e, 124-125, 148
 - parcial total, 184
 - parcial. *Veja* Derivada parcial
 - primeira, 213-216
 - quarta, 218
 - quinta, 218
 - regras de. *Veja* Regras de diferenciação
 - segunda, 217-222
 - terceira, 218
 - total, 181-185, 199-200
 - Derivadas parciais
 - cruzadas (mistas), 281
 - de segunda ordem, 281-282
 - Derivadas totais, 181-185
 - aplicadas à estática comparativa, 199-200
 - parciais, 184
 - Descartes, R., 17
 - Desconto, 253, 269. *Veja também* Valor presente
 - Desconto por quantidade, 127n
 - Desemprego,
 - inflação e, 511-515, 558-562, 586-590
 - política monetária e, 512
 - taxa natural de, 515, 561
 - Desigualdade, 132-135
 - contínua, 132
 - regras de, 132
 - sentido da, 132
 - solução de, 134-135
 - valores absolutos e, 133-134
 - Desigualdade triangular, 65
 - Desvio, 232
 - Determinante, 45, 49, 87-97
 - de n -ésima ordem, 89-92
 - de primeira ordem, 133n
 - de segunda ordem, 87
 - de terceira ordem, 88-89
 - de valor zero, 87-88, 94
 - definição, 87
 - expansão de Laplace de, 89-91
 - fatoração, 93
 - hessiano. *Veja* Determinante hessiano
 - jacobiano. *Veja* Determinante jacobiano
 - nulo, 87, 93-94
 - propriedades de, 92-94, 96
 - Determinante de valor zero, 88, 93
 - Determinante hessiano, 289, 298, 300-301
 - aumentado, 339-344, 352-353, 418n
 - determinante jacobiano em relação a, 325-326
 - Determinante jacobiano, 45
 - em relação a hessiano aumentado, 340
 - em relação a hessiano, 325-326
 - variável endógena, 194, 198, 202, 325-326, 334
 - Determinante nulo, 88, 93
 - Diagonal principal, 55
 - Diagonalização de matriz, 295
 - Diagrama(s). *Veja também* Diagrama de fase
 - de Argand, 491
 - de Venn, 14
 - Diagrama de fase
 - análise de, 627
 - construção de, 626-627
 - estabilidade dinâmica de equilíbrio e, 474-477, 540-543, 594-595
 - para equação de diferenças, 540-544
 - para equação diferencial, 474-477, 479-480
 - para sistema de equações diferenciais, 590-598
 - Diferença
 - primeira, 524
 - segunda, 545

Diferença de vetores, 62
 Diferenciabilidade
 continuidade em relação a,
 139-142
 dupla, 148, 217
 Diferenciação
 diferenciabilidade *versus* 138
 regra da função exponencial de,
 264
 total, 177, 182
 Diferencial total, 176-179, 334
 da função poupança, 177
 de segunda ordem, 282-283,
 286-287, 338
 Diferencial
 regras de, 179-181
 total. *Veja* Diferencial total
 Dinâmica, 425
 de inflação e desemprego,
 511-515, 558-562
 de inflação e regra monetária,
 604
 de investimento, 478-481
 de modelos de insumo-produção,
 580-585
 de preço de mercado, 459-463,
 506-510, 533-540, 543-544
 de renda nacional, 552-557
 do capital, 478-481
 integração e, 425-427
 Discriminação de preço, 317-319
 Discriminante
 aumentado, 339-344
 determinante *versus*, 288
 Distância, 64-65
 Divergência *versus* convergência,
 555-557. *Veja também*
 Convergência
 Domar, E. D., 450n
 Dominação de raízes características,
 551
 Domínio, 19
 Dorfman, R., 45n
 Dualidade, 413n, 415
 Dunn, Sarah, 78n

E

e, o número, 248-249
 Econometria *versus* economia
 matemática, 4-5
 Economia matemática
 definição, 3
 econometria *versus*, 4-5
 economia não-matemática
 versus, 3-4
 Efeito cruzado, 361
 Efeito da escala, 531
 Efeito espelho, 532, 534
 Efeito renda, 360-361
 Efeito substituição, 361

Eixo imaginário, 491
 Elasticidade
 de demanda, 179, 318-319
 de insumo ótimo, 374
 de produção, 368
 parcial, 178, 179
 pontual, 274
 regra da cadeia da, 275
 Eliminação de variáveis, 33-35,
 109, 114
 Empresa multiprodutos, 314-316,
 325
 Enthoven, A. C., 350n, 404n
 Equação(ões)
 auxiliar, 486
 de Bernoulli, 473, 480
 característica. *Veja* Equações
 características
 condicional, 9
 de coestado, 609, 610, 613-614
 cúbica, 36n
 definicional, 8
 diferencial. *Veja* Equação
 diferencial
 exponencial, 255, 258
 homogênea, 456, 458
 de movimento, 607, 609
 não-homogênea, 456-458
 quadrática. *Veja* Equação
 quadrática
 homogênea, 457
 de estado, 609-610, 619
 comportamental, 8-9
 Equações características, 486,
 578-579
 de equação de diferenças, 547
 de equação diferencial, 486
 de sistema de equações de
 diferenças, 572, 575
 de sistema de equações
 diferenciais, 577
 matriciais, 293
 Equações de diferenças simultâneas
 aplicadas, 580-585, 589
 resolvendo, 571-573
 Equações dinâmicas
 de ordem alta, transformação de,
 570-571
 simultâneas, resolvendo, 571-579
 Equações polinomiais
 de grau mais alto, 38-40
 raízes de, 38-40, 519
 Equilíbrio, 31-47
 definição, 31
 economia aberta, 204-206
 em análise de renda nacional,
 46-47
 estabilidade dinâmica de. *Veja*
 Estabilidade dinâmica de
 equilíbrio
 geral, 41-45

intertemporal, 460
 meta, 32, 211
 móvel *versus* constante, 461-462
 parcial, 43
 tipos de, 594-595
 Escala semilogarítmica (semilog),
 272n
 Escalar, 53, 59-60
 Elasticidade pontual, 274
 Espaço de fase, 591
 Espaço de n dimensões, espaço n
 dimensional, 60, 64, 65
 Espaço euclidiano de n dimensões,
 60, 64, 65
 Espaço métrico, 65
 Espaço vetorial, 63-65
 Estabilidade dinâmica de equilíbrio,
 461
 com tempo contínuo, 489-490,
 504-506
 com tempo discreto, 530-533,
 550-552
 diagrama de fase e, 475-477,
 540-542, 594-595
 estabilidade local de sistema
 não-linear, 599, 600-605
 teorema de Routh e, 520-521
 Estabilidade dinâmica, 476
 Estado constante, 480
 Estado estacionário, 480
 Estática, 32. *Veja também* Estática
 comparativa
 Estática comparativa, 118, 121-122
 de empresa multiprodutos, 325
 de modelos de mercado, 196-197
 de modelos de renda nacional,
 200-203
 derivada total aplicada a, 199-200
 do modelo de decisão
 insumo-produção, 325-327
 do modelo de maximização da
 utilidade, 358-362
 modelo de combinação de custo
 mínimo, 372-375
 Estoque, modelo de mercado com,
 537-540
 Excesso de demanda, 32, 41
 ajuste da produção e, 582-584
 ajuste de preço e, 459-460
 em relação ao estoque, 537
 exp, 246
 Expansão de Laplace
 cálculo de um determinante de
 ordem n por, 89-91
 por co-fatores alheios, 97-98
 Expectativas
 adaptativas, 511, 536, 558
 de inflação, 511, 514, 558
 de preço, 506, 536
 Expoente(s), 21, 24, 244
 Exportações líquidas, 203

Extremo, 212
 absoluto *versus* relativo, 212-213, 277, 303, 329
 condição de primeira ordem para, 297
 em relação à concavidade e à convexidade, 302-304
 em relação à quase-concavidade e à quase-convexidade, 353-355
 forte *versus* fraco, 302
 global *versus* local, 212-213
 restrito, 344, 353-355
 teste com determinantes para extremo relativo restrito, 344
 teste com determinantes para extremo relativo, 301
 teste com determinantes para extremo restrito, 344

F

Fator(es)
 de desconto, 253
 de integração, 469
 Fatoração
 de determinante *versus* de matriz, 93
 de função polinomial, 38-39
 de integrando, 431
 Fatorial, 231-232
 Fio da navalha, 452-453
 Flutuação
 amortecida, 504, 539
 degrau, 551-552, 555, 556, 560
 explosiva, 504
 trajetória temporal com, 504-505, 512-515
 uniforme, 504
 Fluxo de caixa, valor presente de, 448
 Fluxo de capital, 203
 Fluxo perpétuo, valor presente de, 449
 Foco, 594
 Folga complementar, 383, 385, 386, 387-388, 398
 Forma, 286
 Forma de Lagrange do resto, 236-237
 Forma linear, 286
 Forma negativa definida, 291
 condições para, 292, 296
 definida *versus* indefinida, 287
 Forma negativa semidefinida, condições para, 296
 definida *versus* semidefinida, 287
 Forma positiva definida, 291
 condições para, 292, 296
 definida *versus* indefinida, 287
 Forma positiva semidefinida
 condições para, 296
 definida *versus* semidefinida, 287

Forma quadrática restrita, 339-340
 Formação de capital, 445-447, 583-584
 Formas quadráticas, 286
 com n variáveis, 292
 condição de sinal definido de teste com determinantes, 287-289
 condição de sinal definido do teste da raiz característica, 292-296
 de três variáveis, 290-292
 restritas, 339-340
 Formby, J. P., 229n
 Fórmula de capitalização, 449
 Fórmula quadrática, 37
 Fração, 9
 Friedman, M., 511
 Função(ões), 18-28
 algébricas *versus* não-algébricas, 22
 argumento de, 19
 circulares, 22, 493-494
 complementares. *Veja* Funções complementares
 côncavas *versus* convexas, 220, 302-304
 constantes, 21, 143-144, 179
 consumo, 46, 553
 continuamente diferenciáveis, 149, 217
 contínuas *versus* descontínuas, 137
 crescentes *versus* decrescentes, 157
 cúbicas, 21, 22, 36n, 38, 227-230
 de Cobb-Douglas. *Veja* Função produção de Cobb-Douglas
 de duas variáveis, valores extremos de, 279-286
 de Lagrange. *Veja* Funções de Lagrange (lagrangeanas)
 definição, 18
 derivadas, 123
 diferenciáveis, 308-310, 349-353
 domínio de, 19, 20
 escalonadas, 127
 exponenciais. *Veja* Função(ões) exponencial
 forma gráfica de, 22
 gerais *versus* específicas, 27-28
 hamiltonianas. *Veja* Função hamiltoniana
 homogêneas. *Veja* Funções homogêneas
 homotéticas, 373-374
 imagem de, 19, 20
 implícitas, 186-189
 inversas, 157, 259, 598
 lineares, 21, 23, 27
 logarítmicas. *Veja* Funções logarítmicas

lucro, 407-408
 objetivo, 212, 298-302, 608, 619
 perda social, 68
 polinomiais. *Veja* Funções polinomiais
 ponto de sela de, 280, 284, 287
 produção. *Veja* Funções produção quadráticas, 21, 23, 27, 35-36
 quase-côncavas *versus* quase-convexas, 345-352
 racionais, 21-23, 138
 série de Taylor de, 599
 senoidais, 493
 transcendentais, 22
 trigonométricas, 22, 493
 valor de, 19, 20
 valor máximo, 406-413
 zeros de, 36
 Função(ões) exponencial(is), 22, 23, 243, 244-254
 base de, 244, 246
 conversão de base de, 261-262
 crescimento e, 248-254
 derivada de, 264-266
 desconto e, 253
 forma gráfica de, 244-245
 funções logarítmicas e, 259
 generalizada, 245-246
 juros compostos e, 250
 neperiano, 246
 série de Maclaurin de, 248
 Funções complementares
 de equação de diferenças de ordem mais alta, 546, 547-550, 571-572
 de equação de diferenças de primeira ordem, 527
 de equação diferencial de coeficiente variável, 464
 de equação diferencial de ordem mais alta, 484-485, 501-503, 519
 de equação diferencial de primeira ordem, 457, 458
 de equações de diferenças simultâneas, 574, 575, 577
 estabilidade dinâmica de equilíbrio e, 460, 530
 Funções côncavas, 313
 critérios de verificação, 304-308
 em programação não-linear, 403
 funções convexas *versus*, 220, 302-304
 Funções continuamente diferenciáveis, 149, 217
 Funções convexas
 conjunto convexo *versus* 310-313
 critério de verificação, 304-308
 em programação não-linear, 403
 funções côncavas *versus*, 220, 302-304

Funções custo, 9
 cúbicas, 227-230
 relação entre marginal e total, 124-125, 148, 444-445
 relação entre média e marginal, 154
 Funções de Lagrange (lagrangeanas)
 em programação não-linear, 382, 388, 389
 para encontrar valores estacionários, 331-334, 336
 Funções de produção linearmente homogêneas, 364-366
 Funções de produção
 CES, 376-378
 de Cobb-Douglas. *Veja* Função de produção de Cobb-Douglas
 função estritamente côncava aplicada a, 324
 função estritamente quase-côncava aplicada a, 371-372
 linearmente homogêneas, 364-366
 Funções demanda hicksianas, 414
 Funções diferenciáveis, 308-310, 349-353
 Funções duplamente continuamente diferenciáveis, 149, 217
 Funções estritamente côncavas, 302-304
 aplicadas a funções de produção, 324
 critérios de verificação, 304-308
 definição, 220
 estritas *versus* não-estritas, 302
 Funções estritamente convexas aplicadas a curvas de indiferença, 356-357
 aplicadas a isoquantas, 324
 critérios de verificação, 304-308
 definição, 220
 estritas *versus* não-estritas, 302-303
 Funções homogêneas
 aplicações econômicas de, 362, 363-369
 linearmente, 363-366, 368-369
 Funções logarítmicas, 22, 23, 259-263
 base de, 254-256
 funções exponenciais e, 259
 Funções polinomiais, 20-21
 continuidade de, 138
 fatoração de, 38-39
 grau de, 21
 limite de, 137
 série de Maclaurin de, 231
 série de Taylor de, 232-233
 Funções valor máximo, 406-413

G

Gamkrelidze, R. V., 609n
 Grau
 de equação diferencial, 455
 de equações polinomiais de grau mais alto, 38-40
 de função polinomial, 21-22

H

Hawkins, D., 114n
 Hipérbole retangular, 21-23, 539, 557
 Hipersuperfície, 26
 Horizonte de tempo infinito, 624-627

I

i, o número, 491
 Identidade, 8
 de Roy, 416, 419
 equilíbrio, 196, 198, 201, 202
 Igualdade
 de conjuntos, 11-12
 de matrizes, 52, 57
 Ilusão monetária, 361
 Imagem, 19, 20. *Veja também* Imagens especulares
 Imagens especulares
 em funções exponenciais e logarítmicas, 259-260
 em hessiano aumentado, 344
 em matriz simétrica, 73
 em trajetórias temporais, 531-533
 Inclinação, 21
 Incremento de renda, 526
 Independência. *Veja* Dependência
 Índice do somatório, 57
 Inflação, 511
 desemprego e, 511-515, 558-562, 586-590
 monetária, 604
 taxa real *versus* taxa esperada de, 544
 Informação qualitativa, 152, 197
 Informação quantitativa, 152, 197
 Instabilidade dinâmica, 476
 Insumo ótimo, elasticidade de, 374
 Insumo primário, 111
 Integração, 426
 constante de, 427
 dinâmica e, 425-426
 limites de, 435, 440, 441-443
 por partes, 433-434, 440
 Integral, 427, 455
 aplicações econômicas de, 444-449
 de Riemann, 438, 439
 de um múltiplo, 431-432
 de uma soma, 430-431

definida, 427, 435-441
 imprópria, 441-444
 indefinida, 427-435, 440
 inferior *versus* superior, 438
 particular. *Veja* Solução particular
 Integrando, 427
 fatoração de, 431
 infinito, 443-444
 Inteiros positivos, 9
 Interseção
 com o eixo horizontal, 260
 com o eixo vertical, 21
 Intervalo aberto, 129
 Intervalo fechado, 129
 Intervalo, fechado, *versus* aberto, 129
 Inverso, 57
 Invertibilidade, 74
 condições para, 83, 94-95
 teste de, 87-92
 Investimento, 201, 450-453
 bruto, 446
 dinâmica de, 477-481
 formação de capital e, 445-447
 induzido, 553
 líquido, 446, 447
 substituição de, 446
 Irregularidades de fronteira, 391-393, 394
 Isocusto, 371
 Isoquanta, 322-324, 370, 372-373

J

Jogo justo, 221

K

Keynes, J. M., 46, 553
 Kuhn, H. W., 381n, 403

L

Lagrange, J. L., 123
 Layson, S., 229n
 Lei associativa
 de operações com conjuntos, 14
 de operações com matrizes, 66-67, 68
 Lei comutativa
 de operações com conjuntos, 14
 de operações com matrizes, 66-67
 Lei de crescimento exponencial, 243
 Lei distributiva
 de operações com conjuntos, 14-15
 de operações com matrizes, 66-67, 69
 Leibniz, G. W., 123
 Lema de Hotelling, 408, 410, 416
 Lema de Shephard, 416-420

Leontief, W. W., 110
 Limite, 125-131
 à esquerda *versus* à direita, 125-127
 cálculo de, 127-128
 de função polinomial, 136-137
 de integração, 435, 440, 441-443
 visão formal de, 128-131
 Linearização. *Veja* Aproximação linear
 Linearização homogênea, 600
 Linha de demarcação, 591, 592
 Linha de fase, 475, 540, 542
 Linhas de fluxo, 593
 Linhas periféricas, 322
 ln, 255
 Log. *Veja* Logaritmo(s)
 Logaritmo(s), 49-50, 245, 248-258
 comum *versus* neperiano, 255
 elasticidade e, 275
 fórmulas de conversão, 258
 regras de, 256-258
 significado de, 254-255
 Lógica literária, 4
 Lógica matemática *versus* literária, 4
 Lógica matemática, 4
 Lucro, maximização de, 224-227

M

Machlup, F., 31n, 425n
 Mapa de canal, 182, 183, 184, 200
 Mapeamento, 18
 Matrizes, 50-59
 adição de, 52-53, 66-67
 características, 293
 co-fator, 98
 como arranjos, 50-51
 de coeficientes, 51
 definição, 51
 diagonais, 68, 72
 diagonalização de, 295
 dimensão de, 51, 54
 divisão de, 57
 elementos de, 51
 escalonadas, 85
 fatoração de, 93
 hessianas, 299
 idempotentes, 70, 72, 77
 identidade, 55, 68, 70
 igualdade, 52, 57
 inversas, 74-77, 97-101
 invertíveis. *Veja* Invertibilidade
 leis de operações com, 66-69
 Leontief, 112-114
 líder *versus* defasagem, 54-55
 multiplicação de, 54-56, 58, 59-60, 68
 multiplicação por escalar, 53
 nulas, 71
 posto de, 84-86, 95-96

quadradas, 51, 87, 94
 simétricas, 73
 singulares, 72, 74
 subtração de, 53, 66
 transição de Markov, 78-79
 transpostas, 72-73
 vetores como, 51-52
 zero, 71
 Maximização da utilidade, 355-363
 estática comparativa de, 358-363
 recursos exauríveis e, 622-623
 tempo de vida, 620-622
 Máximo. *Veja* Extremo
 McShane, E. J., 241n
 Média ponderada, 311
 Média ponderada, 311
 Menor principal, 114-116
 aumentado, 343
 condição de Hawkins-Simon e, 289, 290, 291, 298
 Método do multiplicador de Lagrange, 332-334
 Método dos coeficientes indeterminados, 516-518, 562-565, 581, 584
 Método iterativo para equação de diferenças, 525-527
 Minhas, B. S., 376n
 Minimização de custo, 370-379
 Mínimo. *Veja* Extremo
 Mishchenko, E. F., 609n
 Modelo da teia de aranha, 533-536
 Modelo de crescimento de Domar, 450-453, 454
 Modelo de crescimento de Solow, 477-481, 626
 Modelo de decisão de insumos, 325-327
 Modelo de insumo-produção
 aberto, 111-114
 de Leontief, 110-118
 dinâmico, 580-585
 estático, 110-118
 fechado, 117
 Modelo econômico, 7-9
 Modelo neoclássico de crescimento ótimo, 624-626
 Modelos de insumo-produção de Leontief, 110-118
 Modelos de mercado, 32-44, 105-106
 com estoque, 537-541
 dinâmica de, 459-462, 506-510, 533-539, 543-544
 estática comparativa de, 195-197
 Modelos de renda nacional, 46-47, 106-107
 dinâmica de, 552-557
 equilíbrio na análise de, 46-47
 estática comparativa de, 200-203
 teorema da função implícita aplicado a, 194, 200-203

Modelos e modelagem
 abertos, 111-114
 da teia de aranha, 533-536
 de economia fechada, 107-109
 de mercado. *Veja* Modelos de mercado
 de renda nacional. *Veja* Modelos de renda nacional
 econômicos, 7-9
 fechados, 117
 matemático, 7-9
 Modelos matemáticos, 7-9
 Módulo, 133, 492
 Moeda, utilidade marginal de, 355-356
 Movimento, equação de, 607, 609-610
 Mudança de acordos oficiais, 204n
 Multiplicação por escalar, 53
 Multiplicador
 de Lagrange. *Veja* Multiplicador de Lagrange
 de Lagrange
 interação de, com acelerador, 553-557
 keynesiano, 553
 Multiplicador de Lagrange
 de valor corrente, 619
 interpretação econômica do, 334-335, 355-356, 370
 interpretação geral do, 412-413

N

Nerlove, M., 536
 Neyman, J., 381n
 Nó, 594, 602, 604
 Normalização
 de equação diferencial, 455n
 de vetor característico, 293
 Notação $_$, 57-58
 Notação de conjuntos, 10-11
 Número imaginário, 491
 Número irracional, 10
 Número racional, 10
 Números complexos conjugados, 492
 Números complexos, 491
 conjugados, 492
 expressões alternativas para, 498-500
 Números inteiros, 9
 n -upla ordenada, 51

O

Obst, N. P., 604
 Oferta, 32, 33, 36
 com expectativas de preço, 506
 Ordenada, 37
 Ortante, 350
 Oscilação, 531, 542

explosiva, 543, 573
 trajetória temporal com, 534-536,
 539, 543-544
 Otimização. *Veja também* Extremo
 dinâmica, 421, 607
 problemas de maximização e
 minimização e, 212
 restrita, 411-412
 sem restrição, 406-410
 Ótimo livre, 329
 Ótimo restrito *versus* livre, 329
 Ótimo restrito, 329

P

Par ordenado, 16-17, 18
 Parábola, 21
 Paralelogramo, 62
 Parâmetro, 8
 de distribuição, 376
 de eficiência, 368, 376
 de substituição, 376
 Período, 495, 523
 Phillips, A. W., 511n
 Piso de preço, 544
 Política fiscal, 512
 Política monetária, 512, 558
 Ponto de expansão, 231
 Ponto de inflexão, 215, 220, 223n,
 239-240, 280
 Ponto de sela
 de função, 280, 284, 287
 de sistema dinâmico, 594
 ramos estáveis e instáveis de, 594
 Ponto estacionário, 214
 Ponto terminal fixo, 614
 Pontryagin, L. S., 609n
 Posto, 84-86, 95-96
 Preço de mercado, dinâmica de,
 459-463, 506-510, 533-539,
 543-544
 Preço, trajetória temporal do,
 507-510
 Previsão perfeita, 515
 Primitiva, 427. *Veja também* Integral
 Princípio do máximo, 609-614
 Princípio do máximo de Pontryagin,
 609-614
 Problema primordial, 413
 Problemas autônomos, 619-620
 Problemas duais, 413-420
 Produção de equilíbrio, 225
 Produção ótima, 225
 Produto
 cartesiano, 17
 direto, 17
 escalar, 60, 66
 físico marginal, 157, 189, 323, 478
 interno, 55
 marginal, 322
 receita marginal, 157

Programação côncava, 403
 Programação linear em relação à
 programação não-linear, 381
 Programação não-linear, 337n
 aplicações econômicas de,
 397-402
 em relação à programação linear,
 381
 limitações em, 383-387
 teoremas de suficiência em,
 403-406
 Programação quase-côncava, 404
 Propensão marginal a poupar, 445
 Propensão marginal ao consumo, 46,
 201, 526
 Propriedade de aditividade, 439
 Propriedade de invariância, 362

Q

Qualificação de restrição, 391,
 394-397
 Quociente de diferença, 122-123

R

Radiano, 494
 Raiz latente, 292n
 Raízes
 características. *Veja* Raízes
 características
 complexas, 487-489, 492, 549-550,
 556
 de equação polinomial, 38-40,
 519
 de equação quadrática, 36, 38-40,
 487-489
 dominantes, 551
 reais, 487-489, 547-548, 555,
 559-560
 Raízes características
 condição de sinal definido da
 forma quadrática e, 292-296
 de equação de diferenças,
 547-552
 de equação diferencial, 486-490
 de sistema de equações de
 diferenças, 572
 de sistema de equações
 diferenciais, 576
 dominação de, 551
 estabilidade dinâmica de
 equilíbrio e, 489-490, 504,
 550-552
 Raízes duplas, 488
 Raízes múltiplas, 488
 Raízes reais, 487-489
 distintas, 487-488, 547-548
 repetidas, 488-489, 548, 555,
 559-560
 Recíproca, 57

Recurso exaurível, 621-623
 Regra da cadeia, 155-157, 182, 185,
 275
 Regra da função composta, 156.
Veja também Regra da cadeia
 Regra da função constante,
 143-144, 179
 Regra da função de potência,
 144-147
 de integração, 428
 para encontrar a diferencial
 total, 179
 Regra da função de uma função,
 156. *Veja também* Regra da cadeia
 Regra da função exponencial
 da diferenciação, 264
 da integração, 429
 Regra da função implícita, 188-189,
 192, 367
 Regra da função logarítmica
 da diferenciação, 263-264
 da integração, 429
 Regra da soma/diferença, 147-150
 Regra da substituição, 432-433
 Regra de Cramer, 101-105
 Regra de L'Hôpital, 378, 379
 Regra do produto, 149-151, 179
 Regra do quociente, 153, 179
 Regra monetária, 604
 Regras de diferenciação
 regra da cadeia, 156-157, 182, 185
 regra da função constante,
 143-144, 179
 regra da função exponencial,
 144-147, 179
 regra da função exponencial, 264
 regra da função implícita,
 188-189, 192, 367
 regra da função logarítmica,
 263-264
 regra da soma/diferença,
 147-150, 179
 regra do produto, 149-151, 179
 regra do quociente, 153, 179
 Regras de integração
 integração por partes, 433-434,
 440
 regra de potência, 428
 regra de substituição, 432-433
 regra exponencial, 429
 regra logarítmica, 429
 regras de operações, 429-432
 Relação de Phillips, 511
 com expectativas aumentadas,
 511-512, 558
 de longo prazo, 515, 561-562
 Relação lateral, 330. *Veja também*
 Restrição
 Relação, 17
 Relações de Euler, 496-498
 Renda marginal

inclinação ascendente, 229
versus renda média, 151-153

Renda média
 em relação à demanda, 315-316
versus renda marginal, 151-153

Resto
 forma de Lagrange para o,
 236-237
 símbolo para, 234n

Restrição
 casos de múltiplas restrições, 336,
 344
 de capacidade, 399-402
 de desigualdade, 383-387
 efeitos da, 329-331
 em programação não-linear,
 383-387
 linear, 395-398
 orçamentária, 330, 355, 397-398
 razão, 397-398

Restrição de capacidade, 399-402

Restrição de não-negatividade,
 381-382

Restrição de racionamento, 397-398

Restrição orçamentária, 330, 355,
 397-398

Restrição, 330. *Veja também* Limitação

Restrições de desigualdade, 383-387

Restrições lineares, 395-397

Reta a 45 graus, 540

Reta real, 10

Reta terminal
 horizontal truncada, 615-618
 horizontal, 615
 vertical truncada, 615

Retorno (financeiro), 221

Retornos constantes de escala
 (Constant returns to scale –
 CRTS), 364, 366-367, 370, 376

Retornos de escala
 constantes. *Veja* Retornos
 constantes de escala (CRTS)
 crescentes e decrescentes, 370,
 380

Retornos decrescentes, 228, 478

Risco, atitudes em relação ao,
 221-222

S

Samuelson, P. A., 45n, 520n, 552

Série de Maclaurin, 231
 convergente, 248
 da função seno, 497
 de função co-seno, 497
 de função exponencial, 248
 de função polinomial, 231

Série de potências, 230

Série de Taylor, 230
 com resto, 234
 convergente, 237

de funções polinomiais, 233
 de funções, 599
 extremo relativo e, 238-240

Série. *Veja também* Série de
 Maclaurin; Série de Taylor
 convergência de, 237, 248
 de potência, 230
 infinita, 248, 496-498

Séries infinitas, 248, 496-498

Silverberg, E., 406n

Símbolos
 de operador, 144
 matemáticos, 632-634
 para resto, 234n

Simon, H. A., 114n

Sinal de integral, 427

Sinal de somatório, 57-58

Sistema de equações
 consistência e independência em,
 44-45, 84
 dinâmica de. *Veja* Equações de
 diferenças simultâneas;
 Equações diferenciais
 simultâneas
 homogêneas, 103-104, 117, 572,
 575
 lineares, 49, 76-78, 104

Sistema de números reais, 9-10

Sistemas dinâmicos, gênese de,
 569-571

Smith, W. J., 229n

Solow, R. M., 45n, 376n, 453, 477

Solução, 34
 de desigualdades, 134
 de fronteira *versus* interior, 382
 economicamente
 não-vinculatória, 399
 forma reduzida, 325
 matematicamente vinculatória,
 399
 não-constante, 458
 não-negativa, 114-116
 não-trivial, 104, 577
 resultados para sistemas de
 equações lineares, 104-105
 trivial, 103
 verificação de, 458-459

Solução de fronteira, 382

Solução economicamente
 não-vinculadora, 399

Solução interior, 382

Solução matematicamente
 vinculadora, 399

Solução não-constante, 458

Solução não-negativa, 114-116

Solução não-trivial, 104, 577

Solução particular
 de equação de diferenças de
 ordem mais alta, 546-547
 de equação de diferenças de
 primeira ordem, 528

de equação de diferenças de
 termo variável, 562-565

de equação diferencial de ordem
 mais alta, 484-485

de equação diferencial de
 primeira ordem, 457, 458

de equação diferencial de termo
 variável, 516-518

de equações de diferenças
 simultâneas, 574

de equações diferenciais
 simultâneas, 576

equilíbrio intertemporal e, 461,
 484

Solução particular; Equações
 diferenciais simultâneas
 autônomas, 475
 classificação de, 455-456, 463,
 466, 471-472, 483, 518
 com variáveis separáveis, 472
 definida *versus* solução geral de,
 456
 diagrama de fase para, 475-477,
 479-480
 exatas, 466-469
 grau de, 455
 homogêneas, 456, 458
 não-homogêneas, 456-458
 normalização de, 455n
 reduzidas, 457
 solução particular de, 456

Soluções de forma reduzida, 325

Soma de quadrados, 60, 68

Soma ponderada de quadrados, 68

Somando, 57

Subconjunto, 11

Subconjunto apropriado, 12

Substituição
 elasticidade de, 375, 376
 taxa marginal de, 356
 taxa marginal de, técnica, 190,
 370, 375n

Substitutos, 41, 316, 320, 321

Suen, W., 406n

Superfície, 25
 côncava ou convexa, 346
 hipersuperfície, 26
 utilidade, 357-358

T

Takayama, A., 116n, 350n

Taxa de câmbio fixa, 204

Taxa de crescimento
 encontrando, 272-274
 instantânea, 251-252, 272-274

Taxa de depreciação, 253

Taxa de inflação esperada, 514

Taxa de variação, 122
 instantânea, 123
 proporcional, 272n

Taxa instantânea de crescimento, 251-253, 272-274
 Taxa instantânea de variação, 123
 Taxa marginal de substituição técnica, 370
 elasticidade de substituição e, 375n
 valor absoluto e, 190
 Taxa marginal de substituição, 355-356
 Tempo contínuo, 425
 Tempo discreto, 425
 equações de diferenças e, 523-524
 estabilidade dinâmica de equilíbrio com, 530-533, 550-552
 Teorema da continuidade, 137-138
 Teorema da função implícita, 187, 189n, 190-191, 192
 aplicado a modelos de otimização, 325-327, 334-335, 358
 aplicado a modelos de renda nacional, 194, 200-203
 procedimento de aplicação, 206-207
 Teorema da suficiência de Arrow-Enthoven, 404
 Teorema de De Moivre, 500, 549
 Teorema de Euler, 365-366, 368-369
 Teorema de Pitágoras, 65, 491, 610
 Teorema de Routh, 520-521, 567
 Teorema de Schur, 575
 Teorema de suficiência de Kuhn-Tucker, 403
 Teorema de Taylor, 233
 Teorema de Young, 281, 409, 410
 Teorema do envelope, 406-420
 dedução da identidade de Roy e, 416
 funções valor máximo e, 406-413
 para otimização restrita, 406-412
 para otimização sem restrição, 406-410
 Teorema do limite da soma/diferença, 136
 Teorema do limite do produto, 136
 Teorema do limite do quociente, 136
 Teorema do valor médio, 236
 Teoremas de limite, 135-137
 Teoremas de suficiência, 403-406
 Teoria do controle ótimo, 607-628
 aplicações econômicas de, 620-623
 condições terminais alternativas e, 615-619
 princípio do máximo de Pontryagin em, 609-614
 problemas autônomos em, 619

Teste com determinante para condição de definição de sinal de forma quadrática, 287-289
 para extremo relativo restrito, 344
 para extremo relativo, 301
 Teste da derivada primeira, 213-216
 Teste da derivada segunda, 223, 239
 Teste da n -ésima derivada, 240-241
 Teste de qualificação de restrição, 405
 Timing ótimo, 268-272
 Trajetória de expansão, 372-373
 Trajetória de fase, 593
 Trajetória temporal
 análise do diagrama de fase de. *Veja* Diagrama de fase
 com flutuação degrau, 551-552, 555, 556, 560-561
 com flutuação, 504-505, 512-515
 com oscilação, 535-536, 539, 543-544
 constante, 460, 560
 convergente, 504
 do preço, 507-510
 não-convergente (divergente), 504
 não-oscilatória não-flutuante, 555-556
 tipos de, 475-477, 538, 542-544
 Transformação, 18-19, 570-571
 Transitividade, 132
 Transposta, 72-73
 Tripla ordenada, 17
 Tucker, A. W., 381n, 403

U

Utilidade esperada de jogar, 221
 Utilidade marginal do dinheiro, 356
 Utilização, coeficiente de, 452

V

Valor(es)
 absoluto. *Veja* Valor absoluto crítico, 214
 de equilíbrio, 33
 de função, 19, 20
 de produto marginal, 322
 estacionário, 214, 331-337
 extremo, 212, 279-286
 ótimo, 212
 presente, 253, 447-448, 449
 Valor absoluto
 de números complexos, 492
 desigualdade e, 133
 mais alto, de raiz dominante, 551
 taxa marginal de substituição técnica e, 190

Valores de equilíbrio, 33
 Valores estacionários, 214, 331-336
 Variação, taxa de. *Veja* Taxa de variação
 Variáveis endógenas
 determinante jacobiano de, 194, 198, 202, 325-326, 334
 versus variáveis exógenas, 7-8
 Variáveis exógenas, 7-8
 Variável contínua, 425
 Variável de dimensão n , 292, 336
 Variável(is), 287
 contínua *versus* discreta, 425
 de coestado, 609
 de controle, 607
 de escolha, 212
 de estado, 607, 609
 definição, 7
 dependente *versus* independente, 19
 eliminação de, 34, 109, 114
 endógena *versus* exógena, 7-8
 expoentes como, 244
 Versão de minimização das condições de Kuhn-Tucker, 389
 Vetor(es)
 adição de, 61-62
 característicos, 292, 293
 coluna, 51, 54, 55-56
 combinação convexa de, 311-313
 como matriz(es), 51-52
 interpretação geométrica de, 60-62
 linha, 51, 54, 55-56
 nulo, 61, 62-63
 ortogonais, 294
 ortonormais, 294
 raio, 60
 teste, 393-395
 unitário, 63
 zero, 61, 62-63
 Vetor de dimensão n , 60
 Vizinhança, 129
 Vórtice, 504, 602

W

Walras, L., 43, 45n

Z

Zeros de uma função, 36

